

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 8, 2005, 22:51:59 ; Search time 5280.09 Seconds
(without alignments)
11342.728 Million cell updates/sec

Title: US-09-831-804-1_COPY_720_1955

Perfect score: 1236
Sequence: 1 atgagtcgaagtcgacgaac.....aaacatcagtcattctcga 1236

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 4708233 seqs, 24227607955 residues

Total number of hits satisfying chosen parameters: 9416466

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl1:*
1: gb Da:*
2: gb Htg:*
3: gb In:*
4: gb Om:*
5: gb Ov:*
6: gb Pat:*
7: gb Ph:*
8: gb Pl:*
9: gb Pr:*
10: gb Ro:*
11: gb Sts:*
12: gb Sy:*
13: gb Un:*
14: gb Vi:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	1236	100.0	1239	6	BD274350	BD274350 Candida a
2	1236	100.0	2060	6	BD274349	BD274349 Candida a
3	1232.8	99.7	1239	6	AX488193	AX488193 Sequence
4	581.8	47.1	627	6	ARS50053	ARS50053 Sequence
5	438.4	35.5	110000	8	CR382135_06	Continuation (7 of
6	391	31.6	462	6	ARS50055	ARS50055 Sequence
7	228	18.4	110000	8	CR382122_03	Continuation (4 of
8	227.2	18.4	92211	8	CR380953_09	Continuation (10 o
9	217.4	17.6	1360	8	YSCNOFEAT	M50638 Saccharomyc
10	217.4	17.6	1739	8	YSCPTIITA	M50611 Saccharomyc
11	217.4	17.6	37497	8	YSCP9677	M50611 Saccharomyc
12	199.2	16.1	2133	8	YSCRP026A	M33924 S. cerevisia
13	193	15.6	797	11	CNS06ING	AL400514 T7 end of
14	120.6	9.8	110000	8	AE016818_01	Continuation (2 of
15	101	8.2	169063	9	AC138126	AC138126 Homo sapi
16	101	8.2	177299	9	AC073544	AC073544 Homo sapi
17	99.4	8.0	2132	9	AX714042	AX714042 Sequence
18	99.4	8.0	2132	9	AK056088	AK056088 Homo sapi
19	99.4	8.0	2196	6	CQ726027	CQ726027 Sequence

C	20	99.4	8.0	138627	9	AC099500	AC099500 Homo sapi
C	21	99.4	8.0	169500	2	AC138469	AC138469 Homo sapi
C	22	98.2	7.9	2237	9	AK122869	AK122869 Homo sapi
C	23	97.8	7.9	1203	9	D70831	D70831 Homo sapien
C	24	97.8	7.9	235532	9	AC008739	AC008739 Homo sapi
C	25	97.2	7.9	115955	9	AC011494	AC011494 Homo sapi
C	26	96.6	7.8	2320	6	AR270491	AR270491 Sequence
C	27	96.6	7.8	2330	6	HSU35376	U35376 Human repre
C	28	96.6	7.8	68304	10	BX000432	BX000432 Mouse DNA
C	29	96.6	7.8	106811	9	CR556705	CR556705 Mus muscu
C	30	95	7.7	981	9	HSR2P9	X78932 H. sapiens H
C	31	94.6	7.7	186233	9	AC092329	AC092329 Homo sapi
C	32	94.6	7.7	189317	2	AC024483	AC024483 Homo sapi
C	33	94.2	7.6	292390	2	AC105677	AC105677 Rattus no
C	34	93.6	7.6	1389	9	M27879	M27879 Homo sapien
C	35	93.6	7.6	155439	9	AC008626	AC008626 Homo sapi
C	36	93.4	7.6	199585	10	AC124426	AC124426 Mus muscu
C	37	92.4	7.5	3516	6	CQ726030	CQ726030 Sequence
C	38	92.4	7.5	41153	9	AC016628	AC016628 Homo sapi
C	39	92.4	7.5	158430	2	AC012431	AC012431 Homo sapi
C	40	92	7.4	2873	9	HUMRPLK	M55432 Human Kruep
C	41	92	7.4	2873	11	G28705	G28705 EMS3976 Er
C	42	92	7.4	2873	11	HUMSWS1269	G18281 human chrom
C	43	92	7.4	156835	9	AC010620	AC010620 Homo sapi
C	44	92	7.4	191082	2	AC141066	AC141066 Homo sapi
C	45	92	7.4	203396	9	AC073210	AC073210 Homo sapi

ALIGNMENTS

RESULT 1
BD274350 1239 bp DNA linear PAT 17-JUL-2003
Candida albicans tffiiGene (CatfIIIA) and the coded CATfIIIA
LOCUS
DEFINITION
ACCESSION BD274350 GI:33084118
VERSION BD274350.1 GI:33084118
KEYWORDS
SOURCE JP 2002531068-A/2.
ORGANISM Candida albicans
Candida albicans
Bukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
Saccharomycetales; mitosporic Saccharomycetales; Candida.
REFERENCE
1 (bases 1 to 1239)
Pallier, F.B., Camier, S. and Sentenac, A.
Candida albicans tffiiGene (CatfIIIA) and the coded CATfIIIA
Patent: JP 2002531068-A 2 24-SEP-2002;
JOURNAL AVENTIS PHARMA SA
COMMENT
OS Candida albicans
PN JP 2002531068-A/2
PD 24-SEP-2002
PF 09-NOV-1999 JP 2000581204
PR 10-NOV-1998 FR 98/14147
PI FLORENCE BORDON PALLIER, SYLVIE CAMIER, ANDRE SENTENAC PC
C12N15/09, C12N15/09, A61K39/00, A61K5/00, A61P31/10, C07K14/40, PC
C07K16/14,
PC C12N1/19, C12N1/21, C12P21/02, C12Q1/02, G01N33/15, G01N33/50, PC
G01N33/53,
PC G01N33/566, G01N33/569// (C12N15/09, C12R1.725), C12N15/00, C12N15/
00,
PC (C12N15/00, C12R1.725)
CC Candida albicans tffiiGene (CatfIIIA) and the coded CATfIIIA
CC Candida albicans tffiiGene (CatfIIIA) and the coded CATfIIIA
FH Key protein
FT CDS Location/Qualifiers
1..1239
Location/Qualifiers
/organism="Candida albicans"
/mol_type="genomic DNA"
/db_xref="taxon:5476"

FEATURES
source
ORIGIN
Query Match 100.0%; Score 1236; DB 6; Length 1239;

QY	181	AATGTGATAAGCAATTTTTCAGAAAAATCAATTTGGAAAACACATTTGTATCAATTC	240
Db	900	GATTTGTATTAAGCAATTTTTCAGAAAAATCAATTTGGAAAACATATTGTATCAATTC	959
QY	241	GAATAAAAAAACATTCATTTGTTCAGTGTGTGTGAAGGGGTTAATTCTCGACAACACTTG	300
Db	960	GAATAAAAAAACATTCATTCATTTGTTCAGTGTGTGTGAAGGGGTTAATTCTCGACAACACTTG	1019
QY	301	AAAAGACATGAATACCCATACAAAGTCATTTAAATGTACATTTGAAAATTTGTCAAGAA	360
Db	1020	AAAAGACATGAATACCCATACAAAGTCATTTAAATGTACATTTGAAAATTTGTCAAGAA	1079
QY	361	GCATTTTATTAACATCAATCTTTTAAGACATCATATATTAATCTGTTCAATGAAAAACATTA	420
Db	1080	GCATTTTATTAACATCAATCTTTTAAGACATCATATATTAATCTGTTCAATGAAAAACATTA	1139
QY	421	ACGTGTAAACATGTAATTAAGTTTTCCTCGACCTTCAAAATTAGACACATTAATTA	480
Db	1140	ACGTGTAAACATGTAATTAAGTTTTCCTCGACCTTCAAAATTAGACACATTAATTA	1199
QY	481	AAACATCATGSGTGNATCTCTGCTTATCAATGTGTATCAATCTGTGTTTAAAAATTC	540
Db	1200	AAACATCATGSGTGNATCTCTGCTTATCAATGTGTATCAATCTGTGTTTAAAAATTC	1259
QY	541	CAAACTTGSTGATATTAATTCATTTATTAATAACATCTGATCCAAAATTAAATGTCTT	600
Db	1260	CAAACTTGSTGATATTAATTCATTTATTAATAACATCTGATCCAAAATTAAATGTCTT	1319
QY	601	AAATGTGGTAAAGTGTGTGGGAAAAAGGTTATCTTCACATGTATTAAGTCATGAT	660
Db	1320	AAATGTGGTAAAGTGTGTGGGAAAAAGGTTATCTTCACATGTATTAAGTCATGAT	1379
QY	661	GATTCATCAATGATCAAAATATATGACCTGTGATTTATGTGTGTGGGAAATTTGCAAAAG	720
Db	1380	GATTCATCAATGATCAAAATATATGACCTGTGATTTATGTGTGTGGGAAATTTGCAAAAG	1439
QY	721	AAAAATGAATTAGTTGAACATTAATATATCTTCACATGATGTATATATCTCGATGATTTA	780
Db	1440	AAAAATGAATTAGTTGAACATTAATATATCTTCACATGATGTATATATCTCGATGATTTA	1499
QY	781	TTTAAAGGAACCTGAAGTGAAAAAATTTAGAGAACCTATTAGATCAAGATTCGAATTAAT	840
Db	1500	TTTAAAGGAACCTGAAGTGAAAAAATTTAGAGAACCTATTAGATCAAGATTCGAATTAAT	1559
QY	841	AAATTTGCAATTAAGAAACAGAGAAATTAAGAGTGAAGAAAGATGAAGATGAAGAA	900
Db	1560	AAATTTGCAATTAAGAAACAGAGAAATTAAGAGTGAAGAAAGATGAAGATGAAGAA	1619
QY	901	GATAGCTAGATGAAAAAAGAAAGATGTTATGATTCAGATCAGTCAATGTCACTCAAAAGATCA	960
Db	1620	GATAGCTAGATGAAAAAAGAAAGATGTTATGATTCAGATCAGTCAATGTCACTCAAAAGATCA	1679
QY	961	ATAAAAATCATTTTACTGCTTCTTTGGAGGTTCAAAAGATGTTTCTAAACTTAATTCGAT	1020
Db	1680	ATAAAAATCATTTTACTGCTTCTTTGGAGGTTCAAAAGATGTTTCTAAACTTAATTCGAT	1739
QY	1021	AGTGGGAAGAAAGATCAATTTGTCTTAAGAAATTAATGTGTATAGAAATTTCTTAAAGAAAT	1080
Db	1740	AGTGGGAAGAAAGATCAATTTGTCTTAAGAAATTAATGTGTATAGAAATTTCTTAAAGAAAT	1799
QY	1081	GATTTACGCTGACATTTGAAAAATGTGCATATGATTAATTAACAAGAAATTTGAGTCAATCTTA	1140
Db	1800	GATTTACGCTGACATTTGAAAAATGTGCATATGATTAATTAACAAGAAATTTGAGTCAATCTTA	1859
QY	1141	AATATGTATAGAAAAAGAAAGAACTTCAGAGGTGAACATTTGTTAAAAAAGCCAGATG	1200
Db	1860	AATATGTATAGAAAAAGAAAGAACTTCAGAGGTGAACATTTGTTAAAAAAGCCAGATG	1919
QY	1201	GATTTATTTGCCAAATGAACATCACTGATATTTCTCGA	1236
Db	1920	GATTTATTTGCCAAATGAACATCACTGATATTTCTCGA	1955

RESULT 3				
AX489193				
LOCUS	AX489193	1239 bp	DNA	linear PAT 16-AUG-2002
DEFINITION	Sequence 6493 from Patent WO02053728.			
ACCESSION	AX489193			
VERSION	AX489193.1	GI:22323205		
KEYWORDS	.			
SOURCE	Candida albicans			
ORGANISM	Candida albicans			
	Eukaryotes; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;			
	Saccharomycetales; mitosporic Saccharomycetales; Candida.			
REFERENCE	1.			
AUTHORS	Roemer, T., Jiang, B., Boone, C., Bussey, H. and Ohlsen, K.L.			
TITLE	Gene disruption methodologies for drug target discovery			
JOURNAL	Patent: WO 02053728-A 6493 11-JUL-2002;			
	Elitra Pharmaceuticals, Inc. (US)			
FEATURES	location/Qualifiers			
source	1..1239			
	/organism="Candida albicans"			
	/mol_type="unassigned DNA"			
	/db_xref="taxon:5476"			

Query Match	99.7%	Score 1237.8	DB 6	Length 1239
Best Local Similarity	99.8%	Pred. No. 4.Bc-189		
Matches 1234	Conservative	0	Mismatches 2	Indels 0
				Gaps 0
QY	1	ATGAGTAAAGTACGAAACCAATCCATATCATCTTTAATATCTTCTTCTTCATCA	60	
DB	1	ATGAGTAAAGTACGAAACCAATGCATATCATCTTTAATATCTTCTTCTTCATCA	60	
QY	61	CGTCCCAAAAAGTATATTTGGCATATGAAGGGTGTATTAAGGCTTAAATCGACATCA	120	
DB	61	CGTCCCAAAAAGTATATTTGGCATATGAAGGGGTGTATTAAGGCTTAAATCGACATCA	120	
QY	121	TTATTAGGACACATTAAAGACCAAGTAAATGATGACCGGTAAATAGTACAGTGAAC	180	
DB	121	TTATTAGGACACATTAAAGACCAAGTAAATGATGACCGGTAAATAGTACAGTGAAG	180	
QY	181	GATTGTGATAAAGCATTTTCAGAAATCATATTGGAAACATATTTGATATCATCTCC	240	
DB	181	GATTGTGATAAAGCATTTTCAGAAATCATATTGGAAACATATTTGATATCATCTCC	240	
QY	241	GAIAAAAAAACCATTCATTTGTCAGTGTGTGTGAAGGGTTAATCTCGACACACTTG	300	
DB	241	GAIAAAAAAACCATTCATTTGTCAGTGTGTGTGAAGGGTTAATCTCGACACACTTG	300	
QY	301	AAAAGACATGAATCACCACATACAAAGTCATTTAAATGTCATTTGAAATATGTCAAGAA	360	
DB	301	AAAAGACATGAATCACCACATACAAAGTCATTTAAATGTCATTTGAAATATGTCAAGAA	360	
QY	361	GCATTTTATAAACATCATCTTTAAGACATCATATATTAATCTGTTCAATGAAAAACATTA	420	
DB	361	GCATTTTATAAACATCATCTTTAAGACATCATATATTAATCTGTTCAATGAAAAACATTA	420	
QY	421	ACGTGTAAACATGTATAAAGTTTCACTCGACCTCAAAATTAAGCAACATTAATTA	480	
DB	421	ACGTGTAAACATGTATAAAGTTTCACTCGACCTCAAAATTAAGCAACATTAATTA	480	
QY	481	AAACATCATGTGTGATCTCGCTTATCAATGTATCATCTCGTGTGTTTTTAAAAATTTTC	540	
DB	481	AAACATCATGTGTGATCTCGCTTATCAATGTATCATCTCGTGTGTTTTTAAAAATTTTC	540	
QY	541	CAAACTTGTCAGTATTACAAATTCATATATAAACACTGCATCCAAACTTAAATGTCTCT	600	
DB	541	CAAACTTGTCAGTATTACAAATTCATATATAAACACTGCATCCAAACTTAAATGTCTCT	600	
QY	601	AAATGTGTAAAGGTGTGTGTTGGAAAAAAGTTTATCTTCAACATATGTATAGTCATGAT	660	
DB	601	AAATGTGTAAAGGTGTGTGTTGGAAAAAAGTTTATCTTCAACATATGTATAGTCATGAT	660	
QY	661	GATTCACACATGATCAAAATATGACCTTGATATATGTGATGTGAGTGGGAAATTTGCAAG	720	

Dh 661 GATTTCACATGATCAAAATATGACTTGATTAATGTGAGTGGGAAATTGGCAAG 720
Qy 721 AAAAATGAATTAAGTGAACATTAATATCTTCATGATGATGTAATATCCCTGATGATTA 780
Dh 721 AAAAATGAATTAAGTGAACATTAATATCTTCATGATGATGTAATATCCCTGATGATTA 780
Qy 781 TTTAAGGAAATCTGAGTGAATAAATTAGAACTTAATTAGATCAAGATGCAAAATTAAT 840
Dh 781 TTTAAGGAAATCTGAGTGAATAAATTAGAACTTAATTAGATCAAGATGCAAAATTAAT 840
Qy 841 AATTTCATGAATTTAGAAACAGAGAAATTAAGAGGAAAGATGAAGAGATGAAGAA 900
Dh 841 AATTTCATGAATTTAGAAACAGAGAAATTAAGAGGAAAGATGAAGAGATGAAGAA 900
Qy 901 GATAGCTAGATGAAAAAGAGATGATGATGATGATGATGATGATGATGATGATGATGAT 960
Dh 901 GATAGCTAGATGAAAAAGAGATGATGATGATGATGATGATGATGATGATGATGATGAT 960
Qy 961 ATTAATTCATTTATCTGCTTTTGAAGGTTCAAAGGTTCTAACTTAATCTTAAT 1020
Dh 961 ATTAATTCATTTATCTGCTTTTGAAGGTTCAAAGGTTCTAACTTAATCTTAAT 1020
Qy 1021 AGTGGAGAGAGATCAATGCTCTAAGAAATTAATGATGAATGATGATGATGATGATGAT 1080
Dh 1021 AGTGGAGAGAGATCAATGCTCTAAGAAATTAATGATGAATGATGATGATGATGATGAT 1080
Qy 1081 GATTTCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1140
Dh 1081 GATTTCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1140
Qy 1141 AATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1200
Dh 1141 AATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1200
Qy 1201 GATTTCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1236
Dh 1201 GATTTCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1236

RESULT 4
AR550053
LOCUS AR550053 627 bp DNA linear PAT 08-OCT-2004
DEFINITION Sequence 5184 from patent US 6747137.
ACCESSION AR550053
VERSION AR550053.1 GI:53943228
KEYWORDS
SOURCE Unknown.
ORGANISM Unclassified.
REFERENCE 1 (bases 1 to 627)
AUTHORS Weinstein, K.G. and Bush, D.
TITLE Nucleic acid sequences relating to Candida albicans for diagnostics
and therapeutics
JOURNAL Patent: US 6747137-A 5184 08-JUN-2004;
FEATURES
source
1..627
/organism="unknown"
/mol_type="genomic DNA"

ORIGIN
Query Match 47.1%; Score 581.8; DB 6; Length 627;
Best Local Similarity 97.4%; Pred. No. 3.6e-84;
Matches 605; Conservative 0; Mismatches 7; Indels 9; Gaps 1;
Qy 616 TGTGTTGGGAAAAAGGTTATCTTCATATGTTAAGTCATGATGATGATGATGATGATGATGAT 675
Dh 13 TGTGTTGGGAAAAAGGTTATCTTCATATGTTAAGTCATGATGATGATGATGATGATGATGAT 72
Qy 676 AAAATATGACTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 735
Dh 73 AAAATATGACTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 132
Qy 736 GAACATTATATATCTTCATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 795

Dh 133 GAAACATTAATTAATCTTCATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 192
Qy 796 GTGAAAAATTAAGAACCTATTAATGATCAGAGATGCAATTAATTAATTAATTAATTAATTAAT 855
Dh 193 GTGAAAAATTAAGAACCTATTAATGATCAGAGATGCAATTAATTAATTAATTAATTAATTAAT 252
Qy 856 GAAACAGAAATTAATTAAGTGGAGAGATGAGAGAGATGAGAGAGATGAGAGAGATGAGAGAT 915
Dh 253 GAAACAGAAATTAATTAAGTGGAGAGATGAGAGAGATGAGAGAGATGAGAGAGATGAGAGAT 303
Qy 916 AAAAGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 975
Dh 304 AAAAGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 363
Qy 976 GCTTCCTTGAAGAGTTCGAAAGGTTCTAACTTAATTTCTGAATGATGATGATGATGATGATGAT 1035
Dh 364 GCTTCCTTGAAGAGTTCGAAAGGTTCTAACTTAATTTCTGAATGATGATGATGATGATGATGAT 423
Qy 1036 AATTGCTTAAGAAATTAATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1095
Dh 424 AATTGCTTAAGAAATTAATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 483
Qy 1096 TTGAAATGCGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1155
Dh 484 TTGAAATGCGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 543
Qy 1156 GAAAGAACTCCGAAAGGATGAAACCATTTGTTAAAAAGCAGAGATGATGATGATGATGATGAT 1215
Dh 544 GAAAGAACTCCGAAAGGATGAAACCATTTGTTAAAAAGCAGAGATGATGATGATGATGATGAT 603
Qy 1216 GAAACATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1236
Dh 604 GAAACATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 624

RESULT 5
CR382135_06/c
WPCOMMENT
Sequence split into 16 fragments LOCUS CR382135 Accession CR382135
Fragment Name Begin End
CR382135_00 1 110000
CR382135_01 100001 210000
CR382135_02 200001 310000
CR382135_03 300001 410000
CR382135_04 400001 510000
CR382135_05 500001 610000
CR382135_06 600001 710000
CR382135_07 700001 810000
CR382135_08 800001 910000
CR382135_09 900001 1010000
CR382135_10 1000001 1110000
CR382135_11 1100001 1210000
CR382135_12 1200001 1310000
CR382135_13 1300001 1410000
CR382135_14 1400001 1510000
CR382135_15 1500001 1592360
Continuation (7 of 16) of CR382135 from base 600001 (CR382135 Debaryomyces hanseii chro
Query Match 35.5%; Score 438.4; DB 8; Length 110000;
Best Local Similarity 63.4%; Pred. No. 1.3e-61;
Matches 724; Conservative 0; Mismatches 406; Indels 12; Gaps 3;
Qy 33 ATCTTAATATCTTCTCTTCATCAGTCCCAAAAGTATATTTGACATATGAGAG 92
Dh 20471 ATCTTAATATCTTCTCTTCATCAGTCCCAAAAGTATATTTGACATATGAGAGAG 20412
Qy 93 GTGTATTAAGGCTTAATTCGACATCAATTTTAAAGCAACATTAAGAACCCAGATTA 152
Dh 20411 GTGTATTAAGGCTTAATTCGACATCAATTTTAAAGCAACATTAAGAACCCAGATTA 20352
Qy 153 TGATGACCGTAAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 212

Db	20351	TGAAAGCCCTTTAAATGCTCGGAAACTGGAATCGCAAGATGATTCCTTAAGAAAGTCACA	20292
Qy	213	TTTGGAAACCATATTTGTATC-----ACATTCGGAAAAAAACATTCATGTTCACT	266
Db	20291	TTTACAGCACATTTGGCTTCTCATGAAGATCAAGAAAGTAAACCTTCACATGTGCAAC	20232
Qy	267	GTCGTGTAAGGGGGTTAAATCTCGACAACCTTGAAAGACATGAATCACCCTATACAA	326
Db	20231	TTGGGGCAAAAGGTGAATACTTACCAACTTTAAAGAGACAGAAATATCACATACAA	20172
Qy	327	GTCATTTAAATGTACATTTGAAATTTGTCAAGAGCATTTTATTAACATCAATCTTTAAG	386
Db	20171	ATCATTCATTTGTACATTTAGGGCGGTATGAATCATTTTATTAACATCAATCACTACG	20112
Qy	387	ACATCATATATATCTGTTCATGAATAAAACATTACGTGTAAACAATGTATTAAGTTT	446
Db	20111	ACACCATACCTTATCGGTCATGAATAAAATATTTATGCATATTAAGTATTAATCGTT	20052
Qy	447	CATCGACCTTCAAAATTTAGCAACAATTAATTAAACATCATGTGATCTCCGCTTA	506
Db	20051	TACACAGCCGTATCGATTTGGACACACAATATTAATATCAACAGACTCCCGCTTA	19992
Qy	507	TCAATGTGATCATCTCGTGTGTTTAAATTTTCCAACTTGGTCAGTATTAACAATTTCA	566
Db	19991	TCAATGTGACATCAAGCGCTGTTTGGAAATTTTATGACGTGTCCGCTTTACATTTGCA	19932
Qy	567	TATATAAACATCGATCCAAACTTAATATGCTTAATGTGTAAAGTTGTGTGGGA	626
Db	19931	TATATAAACATGAGACCCAAAGATTAAGTGTCTATATATGTGGAAAGTTGTGGGTGA	19872
Qy	627	AAAAGTTTATCTTACATATGTTAAGTCATGATTTCTACCATGATCAAAATATGAC	686
Db	19871	GAAAGGCTTCGGTCAACATGATATTCACGATGAAGAAAGATGTCAAAATATGGA	19812
Qy	687	TTGTGATTTATGTGAATGGGGAAATTTGCAAGAAAGAAATGAATTGTTGAACATTA	746
Db	19811	TTGTATATATTTGTACATTTGAAAATTTGGAAAAGAAAGCTGATTTATATGATCATTA	19752
Qy	747	TATCTTCATGATGTGTAATATCCCTGATGATTTATTTAAAGSAACTGAAGTGAATAAT	806
Db	19751	CACCTATACAGACAGAAATCTTCTGATGATTTATTTGAACCGATCGAACGAAACAT	19692
Qy	807	AGAGAACCTATTA--GATCAAGATCGAAATTAATATTTGCATGAATTAGAAACAGA	863
Db	19691	AGATTAACCTTCTAAGCGAAGACCTTACCAATATTTGAATTTTAATATCTTTAGAGATT	19632
Qy	864	GAATTTAAAGTGAAGAAAGATGAAGAAAGTGAAGAAAGTATGCTG---ATGAAAAAG	920
Db	19631	ACAATCAAAAGGGTTTGTGAAGTACCGTCGATGAAGAGAAAGAAAGAAATGATTA	19572
Qy	921	AAGATGTTATGATCAGACTCAATGTGAGCTCAAGATCAATTAATCAATTTACTGCTTC	980
Db	19571	GTCGAATGTTTGGACCTTCACTATAATCTTTGAATCTGTGAATTTTAACCTTTGGAATC	19512
Qy	981	TTTGAAGTTCAAGAGTGTCTTAACCTTAATTTTGAATATGTGGAAGAGATCAATTC	1040
Db	19511	TGAAAAAGCATATTTAGATCTTAATTTGAAATAATATCTGAAGAAAGTATCCATG	19452
Qy	1041	TCCCTAAGAAATTTGTGATGAAATGTTTCTAAGAAATATGATTTTACGTGACATTTGA	1100
Db	19451	TCCGAAAAAGAAATGTATGAATGTTCAACCGGTGATTAAGATCTTGAGAGACATTTAA	19392
Qy	1101	ATGCGATGATGATTAATTTTCAAAAGAAATGTAGTCAATCTTAAATAGTATGAAGAAAG	1160
Db	19391	GTCGCAATGATCTCACTTAAGAAAGATTTGAAGCGTTTTAAAGATCTTGAGCTGGAAG	19332
Qy	1161	AA 1162	
Db	19331	AA 19330	

LOCUS	AB550055	462 bp	DNA	linear	PAT 08-OCT-2004
DEFINITION	Sequence 5186 from patent US 6747137.				
ACCESSION	AB550055				
VERSION	AB550055.1	GI:53943230			
KEYWORDS					
SOURCE	Unknown.				
ORGANISM	Unknown.				
REFERENCE	Unclassified.				
AUTHORS	1 (bases 1 to 462)				
TITLE	Weinstock,K.G. and Bush,D.				
JOURNAL	Nucleic acid sequences relating to Candida albicans for diagnostics				
FEATURES	and therapeutics				
FEATURES	Patent: US 6747137-A, 5186-08-JUN-2004;				
SOURCE	Location/Qualifiers				
	1..462				
	/organism="unknown"				
ORIGIN	/mol_type="genomic DNA"				

Query Match	31.6%	Score 391	DB 6	Length 462
Best Local Similarity	98.5%	Pred. No. 2e-53		
Matches 405	Conservative 0	Mismatches 5	Indels 1	Gaps 1

QY	1	ATAGTGAAAGTGCAGAAACCAATGCATATCATCTTTAAATATCTTCTTCTTCA	60
DB	52	ATGATGAAAGTGCAGAAACCAATGCATATCATCTTTAAATATCTTCTTCTTCA	111
QY	61	CGTCCCAAAAAGTATATTTTGCACATAGGAAGGGGTGATAAAGCTATATCGACATCA	120
DB	112	CGTCCCAAAAAGTATATTTTGCACATAGGAAGGGGTGATAAAGCTATATCGACATCA	171
QY	121	TTATTAGGACAACATTTAGAACCCACAGTATATGATCGACGGTATTAATGTACAGTGAAC	180
DB	172	TTATTAGGACAACATTTAGAACCCACAGTATATGATCGACGGTATTAATGTACAGTGAAG	231
QY	181	GATTGTGATTAAGCATTTTTCAGAAATTCACATTTGGAAACACATATTGATCATCTCC	240
DB	232	GATTGTATTAAGCATTTTTCAGAAATTCACATTTGGAAACACATATTGATCATCTCC	291
QY	241	GAAAAAAAACATTCATCTTCAGTGTGTGTGTAAAGGGGTAAATCTTCGACACACTTG	300
DB	292	GAAAAAAAACATTCATCTTCAGTGTGTGTGTAAAGGGGTAAATCTTCGACACACTTG	351
QY	301	AAAAACATGAAATCACCCATACAAAGTCATTTAATGTACATTTGAAAATTTGCAAGAA	360
DB	352	AAAAACATGAAATCACCCATACCAAGTCATTTTAAATGTACATTTGAAAATTTGCAAGAA	411
QY	361	GCATTTTAT-AAAACATCACTTTTAACATCAATATATTTATCTTTCATGA 410	
DB	412	ACATTTTATTAACATCAATCTTTTAACATCAATATATTTATCTTTCATGA 462	

RESULT 7	CR382122_03/c	WPCOMMENT	Sequence split into 14 fragments	LOCUS	CR382122	Accession	CR382122
			Fragment Name	Begin	End		
			CR382122_00	1	110000		
			CR382122_01	100001	210000		
			CR382122_02	200001	310000		
			CR382122_03	300001	410000		
			CR382122_04	400001	510000		
			CR382122_05	500001	610000		
			CR382122_06	600001	710000		
			CR382122_07	700001	810000		
			CR382122_08	800001	910000		
			CR382122_09	900001	1010000		
			CR382122_10	1000001	1110000		
			CR382122_11	1100001	1210000		
			CR382122_12	1200001	1310000		
			CR382122_13	1300001	1320834		
			Continuation of 14 of 14) of CR382122 from base 300001 (CR382122 Kluyveromyces fragilis strain				

[illegible]

	Query Match	Best Local Similarity	Score	Pred. No.	Mismatches	Indels	Gaps
CDS	/gene="P9677.13"	Complement(16937)	17.6%	59.4%	0;	21;	9;
	/gene="P9677.13"	/note="Similar to S. cerevisiae hypothetical protein HRD799 (PIR accession number S45161) and S. pombe					
Db	62 GTCCCAAAAGTATATTGTCACATATGAAAGGGTGATTAAGCCTAATAATGCACCATCAT	121					
Qy	3522 GGCACAAAGACATATTTCTGTGACTATATGCTGTGATTAAGGCAATTTACAAGACCTTCCA	3463					
Db	122 TATTAGCAACATTTAAGAACCCACAGTAATGATCGACCGTAATAATGTAACGTGAGC	181					
Qy	3462 TTTTGACTGAACACCAATTAAGCGTATCAGGGTTTA--AGAGCATTTTCAGTGTGATA	3406					
Db	182 ATTGTGATTAAGCATTTTTAGAAAATCACATTTGGAAACACATATTGTATCATTTCG	241					
Qy	3405 AGTGTGCAAAATCCCTGCTTAAAAAGATCCTTAGAGAGACACTTGTATACGATTCCTG	3346					
Db	242 AAAAAAAAAACATTCATTTGTCACATGATGTGTAAAGGGTAAATTCGACCAACTTGA	301					
Qy	3345 ATAGAAACCATTCCATATGTTCTTATTTGGAAAAGATGACGACTGCCGACACTGA	3286					
Db	302 AAAGACATGAAATCACCCATFACMAAGTCATTTAATGTATCATTTGAAAATTTGTCMAAG	361					
Qy	3285 AGGACACGAAAGTAAAGCATACCAAATCTTTCAATTGTCAGAGAAAGATGCAACCTCC	3226					
Db	362 CATTTTATTAACATCATCTTTAAGACATCATATATTATCTGTTCATGAAAAACATTTAA	421					
Qy	3225 GATTCACAGACATCCCAATTTAAGGCGACATATTTATCTGTTTCATTTACATTAACCTAA	3166					
Db	422 CGGTAAACATGTAATTAAGTTTCTACCTGACCTTGCAAAATTAGCAACATTAATTTAA	481					
Qy	3165 CTGTCCACACTGCAATTAAGAAGCTTTAGAGGCGCATATAGGCTAAGAAATCACATTTCTA	3106					
Db	482 AACATCA-----TGTGTGATCTCTCTCTTATCATGTGATCATCTGTTGTTTAAAA	535					
Qy	3105 AACATCAGATCCCTGAGGTAGAAAAATCTTATCCAAATGATCTTTTGTGTTGTTGCAAG	3046					
Db	536 ATTTCCAAACTTGTGTCAGTATTTACAAATTCATATTAACAACATGATCCAAACTTAAT	595					
Qy	3045 AGTTTCGGAATATGTGTCACAAATTTGCAATTCGATATTAATAATGATCATCTCTAAATTAAT	2986					
Db	596 GTCTTAATATGTGTAAAGGTGTGTGTGGAAAAAGTTTATCTTCACATATGTTAAATGC	655					
Qy	2985 GTCTTAATATGTGTAAAGGTGTGTGTGGAAAAAGTTTATCTTCACATATGTTAAATGC	2926					
Db	656 ATGATGATTTTACACATGATCCAAATATGGAAGCTTGATTAATGATGATGTGGGAAATTTG	715					
Qy	2295 ATGACGACTCTATGTAACAAAAATTGGAAGTGTCTATATATGTCCTGATATGTCITTTT	2866					
Db	716 CAAGAAAAATGAATTAAGTGAACATTAT	744					
Qy	2865 CTAGAAAAATGATCTTCTCAGCATTAAT	2837					
RESULT 12	YSCRPO26A/c	2133 bp DNA linear	PLN 27-APR-1993				
LOCUS	S.cerevisiae RNA polymerase II sixth subunit (RP026) gene, complete						
DEFINITION	cde						
VERSION	MJ3924						
KEYWORDS	RNA polymerase II; c-myc proto-oncogene; transmembrane protein; tyrosine kinase.						
SOURCE	Saccharomyces cerevisiae (baker's yeast)						
ORGANISM	Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes; Saccharomycetales; Saccharomycetaceae; Saccharomycetes.						
REFERENCE	1 (bases 1 to 2133)						

AUTHORS Archambault, J., Schappert, K.T. and Priesen, J.D.
 TITLE A suppressor of an RNA polymerase II mutation of *Saccharomyces cerevisiae* encodes a subunit common to RNA polymerases I, II, and III
 JOURNAL Mol. Cell. Biol. 10 (12), 6123-6131 (1990)
 MEDLINE 91061718
 PUBMED 2247052
 COMMENT Original source text: *S.cerevisiae* single-copy plasmid DNA, clone 2.
 Draft entry and computer-readable sequence for [unpublished (1990)] kindly submitted by J.D.Priesen, 01-MAY-1990
 The Hospital for Sick Children
 555 University Avenue
 Toronto, Ontario, CANADA M5G 1X8
 e-mail: jim@ickids.toronto.edu.
 Location/Qualifiers
 1..2133
 /organism="Saccharomyces cerevisiae"
 /mol_type="genomic DNA"
 /db_xref="taxon:4932"
 join(1026..1045,1122..1569)
 /note="RNA polymerase II sixth subunit (RP026)"
 /codon_start=1
 /protein_id="AAA34989.1"
 /db_xref="GI:172453"
 /translation="MSDYEAFFNDNGENFEDVEHFSDETYEKKPKQGETTIDAN
 GRTIVGNGPDPFOHBOIRRTLEKKAIPDQRAATPKDQATPKREARILIGRAIQIS
 NNAIPVDEEGEDPLRIKMKELAEKIPVIRRIIPDSFEDMSVEELIVL"
 <1026..1045
 /note="RNA polymerase II sixth subunit (RP026)"
 /number=1
 1046..1121
 /note="RP026 intron A"
 1122..1569
 /note="RNA polymerase II sixth subunit (RP026)"
 /number=2
 ORIGIN Chromosome XVI.
 Query Match 16.1%; Score 199.2; DB 8; Length 2133;
 Best Local Similarity 61.4%; Pred. No. 1.1e-22;
 Matches 340; Conservative 0; Mismatches 208; Indels 6; Gaps 1;
 QY 171 TACAGTGAAGATGATTAAGCATTTTTCAGAAAATACATTTGGAAACACATATTGT 230
 DB 556 TTCAAGTGATATGTGGCAAAATCCTTCGTTAAAGAGTCACTTAGAGACACTGTA 497
 QY 231 ATCAGCTTCGCAAAAAAATCCATTCATTTGAGTGTGTGAAGGGGTTAATTCCTG 290
 DB 496 TACGCAATCTGATACGAAACCAATGTTCTTATTTGTGAAAAGAGTACGACTCG 437
 QY 291 ACAACCTTGAAGAGCATGAATACCCGATCAAAAGTATTAAATGATCAATTTGAAA 350
 DB 436 CAGCACTGAAGCGACAGAGTAACGATACCAATTTTCATTTGTCAGAGAAAG 377
 QY 351 TTGTCAAGAGCATTTTATTAACATCTTAAGACATATATATATTCCTGTCATGA 410
 DB 376 ATGCAACCTCCGATTTTACAGATCCACATTTAAGGCAATATTTTCTGTTTCATT 317
 QY 411 AAAACATTACGTTAAACATGTATAAAGTTTCACTCGACCTTCAAAATTAGACA 470
 DB 316 ACATAACTAACCTGTCACACATGCAATAAAGCTTCAGAGGCCATATAGCTAAGAA 257
 QY 471 ACATAAATTAACAACA-----TGGTGATCTCTCGTTATCAATGTGATCAATCC 524
 DB 256 TCAATTTCTTAACATCAAGATCTGAGTAGAAAAATCCTTACCAATGTACTTTGCTG 197
 QY 525 TTGTTTAAAAAATTTCCAACTGTGTCAGTATTAACAATTTCAATATAAACAATGATCC 584
 DB 196 TTGTTGCAAGAGTTTCGATATGTGTCACATTTGCAATGCAATATAAAGATATCC 137
 QY 585 AAACTAAATGTCTTAAATGTGTGAAGGTTGTGTGGAAAAAAGTTTATCTTCACA 644

DB 136 TAAATTAAGTCTTATTGTCAGCAACATGTGTGGGAAAAATGGTTACAATGCA 77
 QY 645 TATGTTAAGTCATGATGATTTTACCATGATCAAAATATGACTTGTGATTTATGATGT 704
 DB 76 CATGATTTATTCATGACGACTACTAGTAAACCAAAATGGAAGTCAATATATGTCCTGA 17
 QY 705 GGGGAATTTGCAA 718
 DB 16 TATGCTTTTCTTA 3
 RESULT 13
 CNS06ING 797 bp DNA linear STS 30-NOV-2001
 LOCUS T7 end of clone AS0AA022B07 of library AS0AA from strain CLIB 533
 DEFINITION of *Saccharomyces bayanus*, sequence tagged site.
 ACCESSION AL400514.1 GI:12156635
 VERSION AL400514.1
 KEYWORDS STS.
 ORGANISM *Saccharomyces bayanus*
 SOURCE *Saccharomyces bayanus*
 BUKARYOTA; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 Saccharomycetales; Saccharomycetaceae; Saccharomyces.
 REFERENCE 1 (bases 1 to 797)
 Boulet, J.L., Aigle, M., Artiguenave, F., Blandin, G.,
 Boulet, J.L., Aigle, M., Bon, E., Brotier, P., Casaregola, S.,
 de-Montigny, J., Dujon, B., Durrens, P., Ledingle, A., Llorente, B.,
 Malberuy, A., Neuvéglise, C., Ozier-Kalogeropoulos, O., Poirier, S.,
 Saurin, W., Tekala, F., Toffano-Nioche, C., Wesolowski-Louvel, M.,
 Wincker, P. and Weissenbach, J.
 Genomic exploration of the hemiascomycetous yeasts: 1. A set of
 yeast species for molecular evolution studies
 FEMS Lett. 487 (1), 3-12 (2000)
 TITLE Yeast species for molecular evolution studies
 JOURNAL FEMS Lett. 487 (1), 3-12 (2000)
 MEDLINE 20584711
 PUBMED 11152876
 REFERENCE 2 (bases 1 to 797)
 Bon, E., Neuvéglise, C., Casaregola, S., Artiguenave, F., Wincker, P.,
 Aigle, M. and Durrens, P.
 Genomic exploration of the hemiascomycetous yeasts: 5.
Saccharomyces bayanus var. *uvatum*
 FEMS Lett. 487 (1), 37-41 (2000)
 PUBMED 20584715
 TITLE Direct Submission
 JOURNAL Submitted (07-SEP-2000) Genoscope - Centre National de Séquençage,
 2 rue Gaston Crémieux, CP 5706, 91057 EVRY cedex, FRANCE. (E-mail :
 seqref@genoscope.cns.fr - Web : www.genoscope.cns.fr)
 This STS is part of a random genomic sequencing program of thirteen
 yeast species: *Saccharomyces bayanus* var. *uvatum*, *Saccharomyces*
exiguus, *Saccharomyces servazzii*, *Zygosaccharomyces rouxii*,
Saccharomyces kluyveri, *Kluyveromyces thermotolerans*, *Kluyveromyces*
lactis var. *lactis*, *Kluyveromyces marxianus* var. *marxianus*, *Pichia*
angusta, *Debaryomyces hansenii* var. *hansenii*, *Pichia sodicophila*,
Candida tropicalis and *Yarrowia lipolytica*. Genomic inserts of 3 to
 5 kb were prepared and both extremities were sequenced. See
 keywords for description of this sequence and for the sequence of
 the other extremity of this insert.
 Location/Qualifiers
 1..797
 /organism="Saccharomyces bayanus"
 /mol_type="genomic DNA"
 /strain="CLIB 533"
 /variety="uvatum"
 /db_xref="taxon:4931"
 /clone="AS0AA022B07"
 /clone_1ib="AS0AA"
 /note="end : 77"
 <2..>796
 /note="similar to *Saccharomyces cerevisiae* ORF YPR186c [PZF1 / TFI1A (transcription initiation factor)]"
 /evidence="not_experimental"

ORIGIN

Query Match 15.6%; Score 193; DB 11; Length 797;

Best Local Similarity 59.3%; Pred. No. 1.3e-21;

Matches 347; Conservative 1; Mismatches 231; Indels 6; Gaps 1;

QY 178 GACGATTTGATTAAGCAATTTTCAGAAAATCATTGGAAAACATATTTGATCAT 237
 DB 83 GACATTTGTCAAAATCTTTGTTAAABAAGATCATTAAGAGGACCTGTTTGGCAT 142
 QY 238 TCCGAAAAAAACCATTTCCATTGTTCACTGTGTGTGTAAGGGGTAAATTTCCACAAC 297
 DB 143 TCTGATACGAAGGCATTTCCATGTTCTTATTTGTGGCAAGGGGTGAACAATCCACAGCA 202
 QY 298 TTGAAAGACATGAATCAACCATACCAAGTCATTTAAATGATCATTTGAAATTTGCA 357
 DB 203 TTGAAGGCCACAGAAAGTTACACAAATCTTTGTTGTCCGGAAGATGATGAT 262
 QY 358 GAAGCATTTTAAACATCATCTTTAAAGACATCATATTTATCTGTTCAATGAAAAACA 417
 DB 263 CTCGATTTTAAAGCACCCTCAATTAAGGGCTCATATTTCTATCTGTCCACCTAGACAA 322
 QY 418 TTAACTGTAAACATGTAATTAAGTTTCACTGCACCTTCAAAATTAAGCAACATMAA 477
 DB 323 TTAACTGTCCACAATGTAGTAAATTTTCAAGAGCCCTAAGGCTAAGAAATCATATC 382
 QY 478 TTAAACATCATGTT-----GGATCTCTCTTATCATATGATCATCTGCTGTTT 531
 DB 383 TCTAAACATCATATCTCTGAAGGTGTAATCCGATCAATGTAATCTTCTGCTGCTGTT 442
 QY 532 AAAAATTTCCAACTTGTGATTAATTAATTTCAATTAATTAATTAATTAATTAATTAAT 591
 DB 443 ACAGAGTACCTGTAATGTGACAGTTGATCATATTAATTAATTAATTAATTAATTAAT 502
 QY 592 AAATGCTTAATGTGTAAGGTGTGTGGGAAAAAGTTTATCTTCAATATGTA 651
 DB 503 AAATGCCCCATTGCAACAGCCTTGTGTAGGGGAAATGCTTCAAAATGATATGATC 562
 QY 652 AGCATGATGATTTTACATGATCAAAATATGACTGTGATTAATTTGTATGTGTGGGAAA 711
 DB 563 ATTCAATGATGAATCTTTGTGTAAGAACTGGAAGTCAATTTGTTCTTAATATCATCT 622
 QY 712 TTGCAAGAAAAATGAAATGATTAGTTGAACATTTATATTTCTTCAT 756
 DB 623 TTCTTGAGAAAGCAGACCTCTTACTACTACCAAGTACGCAAT 667

RESULT 14
 AB016818.01

WPCOMMENT

Sequence split into 16 fragments LOCUS AB016818 Accession AB016818

Fragment Name	Begin	End
AB016818_00	1	110000
AB016818_01	100001	210000
AB016818_02	200001	310000
AB016818_03	300001	410000
AB016818_04	400001	510000
AB016818_05	500001	610000
AB016818_06	600001	710000
AB016818_07	700001	810000
AB016818_08	800001	910000
AB016818_09	900001	1010000
AB016818_10	1000001	1110000
AB016818_11	1100001	1210000
AB016818_12	1200001	1310000
AB016818_13	1300001	1410000
AB016818_14	1400001	1510000
AB016818_15	1500001	1519138

Continuation (2 of 16) of AB016818 from Base 100001 (AB016818 Eremothecium gossypii ATCC
 Query Match 9.8%; Score 120.6; DB 8; Length 110000;
 Best Local Similarity 50.1%; Pred. No. 1.9e-10;
 Matches 386; Conservative 0; Mismatches 374; Indels 11; Gaps 3;

QY 25 TCGATATCATCTTTAATATCTTCTTCTCATCACTGCCAAAAAGTATATTGCACA 84
 DB 17349 TCGTTTCCAGTGTTAAGTTCACAGCGGTGAGAGACCGCGCAAAACGATTACTCGAC 17408
 QY 85 TATGAAGGTGTGATTAAGCCTTAATTAATGACATCATTTATTAAGACAACA-TTTAAGAAC 143
 DB 17409 TACGAGGGCTGCTACAAAGCGTTCACGCGCGGTCCCTGCTGAGAGACACGAGACAG 17468
 QY 144 CCACAGTAATGATCGACCGTATTAATGTAAGTGAAGTGAATGATTAAGCATTTTTCAG 203
 DB 17469 GCGCACACAGGAGATCGAGGCAATACAGTGC---GAGCAGTGGCGCGGTTTACAA 17524
 QY 204 AAAATCACATTTGGAAAACAATATTGATCACTTCGAAAAAAACATTTCCATTGTC 263
 DB 17525 GAAATGCACTTGGACCGGCACTGTTCTGCACTTGAGACAAACCGTTGCTGCTGCAC 17584
 QY 264 AGTGTGTGTAAGGGGTTAATTTCTGACAAACATTTGAAAGAATGAAATCAACCATAC 323
 DB 17585 GGTATGTGCAAGGGGGTGCACCGCGGACGAATGAGGGCGGACAGATTCACACAC 17644
 QY 324 AAGTCAATTTAATGTAATTTGAATTTGAATTTGCAAGAGCATTTTATTAACATCATCTT 383
 DB 17645 AAGTGTTCAGATGCGCCGACAGAGGCTGCGCGGAAGGTTCTTACAAAGCACCGCAGTT 17704
 QY 384 AAGACATCATATTTATTTCTGTTGATGAAAAAATTAATGATTAATTAAGT 443
 DB 17705 GCGTTGCAAGTGTGCGGCTGCGGCTGACGAGCAAAAGCTGAGCTGACACATGTAACAGCG 17764
 QY 444 TTTCATCGACCTTCAAAATTAAGCAACATTAATTAATTAATTAATTAATTAATTAAT 500
 DB 17765 GTTCCAAAGGCGCATTAACGTTTGAACACACATCGGCAAGCATCAAGGGGCTGCATCGCA 17824
 QY 501 ---TGCTTATCAATGTGATCATCTGTTGTTTAAATTTCAAACTTGTGCTAGTAT 557
 DB 17825 GTTCCGATCAATGACAGAACCGCGCTGCTGATGATGATGATGATGATGATGATGAT 17884
 QY 558 ACAATTTCAATTAATTAACATGATCAAACTTAATGCTTAATGCTTAATGCTTAATGCT 617
 DB 17885 GCGACGACCTTCAACAGGACCAACCGAAGCTACCAATGCGGGGTGTGCGGAACTGTG 17944
 QY 618 TGTGGGAAAAAAGTTTATCTTCAATTAATTTAATGATGATGATGATGATGATGATGAT 677
 DB 17945 CGTGGGGAGAGCGGAGCTGAGATGACATGATGATGATGATGATGATGATGATGAT 18004
 QY 678 AATATGACTTGTGATTAATTTGATGATGATGATGATGATGATGATGATGATGATGAT 737
 DB 18005 AAATGGAAGTGTCTGTGTGTGATGACACACACCTACGCGGAAGATGCGCATCTGCT 18064
 QY 738 ACATTTAATATCTTCAATGATGATTAATTTCCCTGATGATTTATTAAGGA 788
 DB 18065 GCATTAATGAAGAAACCCAGGTGATTCATCCCAAGAAATGATTAAGGA 18115

RESULT 15

AC138126/c

LOCUS AC138126 169063 bp DNA linear PRI 17-DEC-2002

DEFINITION Homo sapiens chromosome 19 clone RP11-274A19, complete sequence.

AC138126

VERSION AC138126.1 GI:2751357

KEYWORDS

SOURCE Homo sapiens (human)

ORGANISM

Homo sapiens; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Eukaryota; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 169063)

DOE Joint Genome Institute and Stanford Human Genome Center.

TITLE Direct Submission

JOURNAL Unpublished

2 (bases 1 to 169063)

DOE Joint Genome Institute and Stanford Human Genome Center.

AUTHORS Direct Submission

TITLE Submitted (17-DEC-2002) DOE Joint Genome Institute, 2800 Mitchell

JOURNAL

COMMENT

Drive, Walnut Creek, CA 94598, USA
Draft Sequence Produced by DOE Joint Genome Institute
www.jgi.doe.gov
Finishing Completed at Stanford Human Genome Center
www-shgc.stanford.edu
Quality: Phrap Quality >=40 100% of Sequence;
Estimated Total Number of Errors is 0.

FEATURES

source

1.169063
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="19"
/clone="RP11-274A19"

ORIGIN

Query Match 8.2%; Score 101; DB 9; Length 169063;
Best Local Similarity 50.8%; Pred.No.2.4e-07;
Matches 298; Conservative 0; Mismatches 280; Indels 9; Gaps 2;

QY 7 GAAAGTGAGAAACCAATCGATCATCTTTAATATCTTCTTCTTCATCAGTCCC 66
DB 80134 GAATGGGCAAAAGCTTTTAAACAGTCTCAATCCTTACTACACATAGAAATTCATCT 80075
QY 67 AAAAAGTATTTGCACATATGAGGGTGTGATAAAGCTATATGACCATCATATTATTA 126
DB 80074 GGAGGAAATCCTACAAATGTGAAAGATGGCAAGCTTCTATGATCTCAAAACTT 80015
QY 127 GAGCAACATTTAAGAACCCAGTATGATGACCGTATAAATGTACAGTGAAGATTGT 186
DB 80014 ACTGAACATTAAGAAATTTCACTGAGAGAAACCTTACACAGT-----GAAAGATGT 79961
QY 187 GATTAAGCATTTTTCAGAAATACATTTGAAACACATATTTGATCATCTCCGAAAAA 246
DB 79960 GGCAGAAAGCTTTAACATCTTCCACATCTGTGACACATAGATATTCATCTGAGAG 79901
QY 247 AAACCATTCATTTGCTGAGTGTGTGTAAGGGGTTAATTCGACACACTTGAAAAGA 306
DB 79900 AAACCTTACAAATGTGAAGAAATGTGTAAGCTTTAACAGTCTTCAACCTTACTAGA 79841
QY 307 CATGAATGACCATCAATCAAGTCATTTAATGTACATTTGAAATTTGCAAGAGCATTT 366
DB 79840 CATTAAGAAATTCATCTGAGAGAAACCTTACCAATGTGAAAAAATGTGCAAAAGCTTTT 79781
QY 367 TATTAACATCAATCTTTAA---GACATCATATATTATCTGTCAAGAAAAACATTAACG 423
DB 79780 AACAGTCTCAAACTTACTGAGACATAGAAAAATTCATCTGTGAGAAACTTACAAA 79721
QY 424 TGTAAACAATGTATTAAGTTTCACTGACCTTCAAAATTAGCACACATTAATTAAAA 483
DB 79720 CCTAAAGATGTACAGTATTTTGAACACTTCAAAGTTTCTTAAACATTAAGAAAT 79661
QY 484 CATCATGTGATCTCCCTGCTATCATGTGATCATCCGTGTTTAAATTTCCAA 543
DB 79660 TATGCTGGTGAAGAAATCTTGAAGATGTGAAGATGTACAAAACCTTAAAGTTGTCA 79601
QY 544 ACTTGTCAGTATTAACAATTCATTAATAAACAATGATCCAAACT 590
DB 79600 ACTTGATTTGTGATAGATTAATTCATTAATTAATTAATTAATTAATTAATTAATTA 79554

Search completed: February 9, 2005, 05:36:20
Job time : 5288.09 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 8, 2005, 22:43:04 ; Search time 640.722 Seconds
(without alignments)
11419.610 Million cell updates/sec

Title: US-09-831-804-1_COPY_720_1955

Perfect score: 1236
Sequence: 1 atcgatggaagtcgacgaac.....aaacatcagtgattctcga 1236

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 4390206 seqs, 2959870667 residues

Total number of hits satisfying chosen parameters: 8780412

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_16Dec04:*
1: geneseqn1980s:*
2: geneseqn1980s:*
3: geneseqn2000s:*
4: geneseqn2001as:*
5: geneseqn2001bs:*
6: geneseqn2002as:*
7: geneseqn2002bs:*
8: geneseqn2003as:*
9: geneseqn2003bs:*
10: geneseqn2003cs:*
11: geneseqn2003ds:*
12: geneseqn2004as:*
13: geneseqn2004bs:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	1236	100.0	2060	3	AAA15398	AAA15398 DNA encod
2	1232.8	99.7	1239	6	ABZ32206	ABZ32206 Candida a
3	99.4	8.0	2132	10	ADA53158	ADA53158 Human cod
4	96.8	7.8	1705	9	ACA98970	ACA98970 cDNA enco
5	96.6	7.8	2320	10	ACA56456	ACA56456 Human sig
6	96.6	7.8	2320	12	AD156252	AD156252 Human pol
7	95.2	7.7	4563	5	AA591317	AA591317 DNA encod
8	95	7.7	981	12	ADN04717	ADN04717 Antipsoi
9	94.6	7.7	2597	8	AA151569	AA151569 Human nuc
10	92.4	7.5	3639	5	AA564586	AA564586 DNA encod
11	92	7.4	994	12	ACH87568	ACH87568 Human gen
12	92	7.4	2873	13	ADR14368	ADR14368 Human NF-
13	91.8	7.4	1533	12	ADN99064	ADN99064 Novel hum
14	91.8	7.4	1533	12	ADN00633	ADN00633 Novel hum
15	91.8	7.4	2110	10	ADDB6324	ADDB6324 Human cDN
16	91.8	7.4	2760	10	ADCS6695	ADCS6695 Human mac
17	91.8	7.4	3078	10	ADBE2468	ADBE2468 Human cDN
18	91	7.4	3755	13	ADR07315	ADR07315 Full leng
19	91	7.4	4209	13	ADR08223	ADR08223 Full leng
20	90.4	7.3	2026	8	AAD55863	AAD55863 Human nuc

21	90.4	7.3	2114	10	ADA53124	ADA53124 Human cod
22	90.2	7.3	2230	10	ADA52931	ADA52931 Human cod
23	89.8	7.3	590	12	ACH80103	ACH80103 Human gen
24	89.2	7.2	2064	10	ADC30762	ADC30762 Human nov
25	89.2	7.2	2221	11	ADM01344	ADM01344 Human cDN
26	89.2	7.2	2632	5	AA568872	AA568872 DNA encod
27	89.2	7.2	2729	4	AAH16178	AAH16178 Human cDN
28	89.2	7.2	3839	6	ABK83826	ABK83826 Human cDN
29	89.2	7.2	3839	12	ADN04207	ADN04207 Antipsoi
30	89.2	7.2	3839	13	ADP25049	ADP25049 Breast ca
31	89	7.2	1991	13	AD509824	AD509824 Human the
32	89	7.2	2298	4	AAH16608	AAH16608 Human cDN
33	89	7.2	2298	13	AD509825	AD509825 Human the
34	89	7.2	2905	5	AA566143	AA566143 DNA encod
35	89	7.2	3003	12	ADN06033	ADN06033 Antipsoi
36	89	7.2	3020	5	AA592560	AA592560 DNA encod
37	89	7.2	3502	4	AA157845	AA157845 Human pol
38	88.6	7.2	1335	12	ADN98906	ADN98906 Novel hum
39	88.6	7.2	1335	12	AD000475	AD000475 Novel hum
40	88.6	7.2	1466	12	ACH91879	ACH91879 Human gen
41	88.6	7.2	1875	10	AD160591	AD160591 Secreted
42	88.6	7.2	2597	8	AAD55855	AAD55855 Human nuc
43	88.6	7.2	4227	9	ACA98938	ACA98938 cDNA enco
44	88.4	7.2	1410	12	ACH87337	ACH87337 Human gen
45	88.4	7.2	1952	10	ADC58104	ADC58104 Zinc fing

ALIGNMENTS

RESULT 1	
AAA15398	
ID	AAA15398 standard; DNA; 2060 BP.
AC	AAA15398;
DT	04-SEP-2000 (first entry)
DE	DNA encoding a transcription factor designated CATFIIIA.
KW	Transcription factor; CATFIIIA; DNA-binding protein;
KW	ribosomal RNA 5S gene; fungal infection; ss.
OS	Candida albicans.
XX	
XX	
FT	Key
FT	CDS
FT	Location/Qualifiers
FT	720..1958
FT	/*tag= a
FT	/transl_except= (pos: 1296..1298, aa: Ser)
FT	/transl_except= (pos: 1734..1736, aa: Ser)
PN	WO200028037-A1.
XX	
XX	
PD	18-MAY-2000.
XX	
XX	
PF	09-NOV-1999; 99WO-FR002739.
XX	
PR	10-NOV-1998; 98FR-00014147.
XX	
PA	(HMRI) HOECHST MARION ROUSSEL.
XX	
PI	Bordon-Pallier F, Camier S, Sentenac A;
XX	
DR	WPI; 2000-376549/32.
XX	
DR	P-PSDB; AAY93316.
PT	New nucleic acid encoding Candida albicans transcription factor, useful
PT	e.g. in screening for antimycotic agents and for immunization.
XX	
PS	Claim 4; Page 32-33; 45pp; French.
XX	
CC	The present sequence encodes a Candida albicans transcription factor, designated CATFIIIA. The polypeptide is a DNA-binding protein, which is

CC involved in initiating transcription of the ribosomal RNA 5S gene. The
CC polynucleotide is used to screen for its specific inhibitors, potentially
CC useful as antimycotic agents, to raise an antibody response that is
CC protective against fungal infection and to raise antibodies. Such
CC antibodies, as well as the polypeptides and polynucleotides are used in
CC compositions for diagnosing and treating fungal infections, e.g. by
CC detecting polymorphisms and mutations

XX Sequence 2060 BP; 726 A; 315 C; 354 G; 665 T; 0 U; 0 Other;

Query Match 100.0%; Score 1236; DB 3; Length 2060;

Best Local Similarity 100.0%; Pred. No. 1.4e-233; Mismatches 0; Indels 0; Gaps 0;

Matches 1236; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGAGTGAAGTGAAGCAAAACCAATGATATCATCTTAAATATCTTCTTCTTATCA 60
DB ATGAGTGAAGTGAAGCAAAACCAATGATATCATCTTAAATATCTTCTTCTTATCA 779
QY 61 CGTCCCAAAAAGTATATTTTGACATATGAAAGGTGATTAAGCCATTAATGACCATCA 120
DB CGTCCCAAAAAGTATATTTTGACATATGAAAGGTGATTAAGCCATTAATGACCATCA 839
QY 121 TTATTAGAGCAATTTAAGAACCCACAGTAATGATGACCCGATTAATGTAAGAGGAC 180
DB TTATTAGAGCAATTTAAGAACCCACAGTAATGATGACCCGATTAATGTAAGAGGAC 899
QY 181 GATTGTGATTAAGCATTTTTCAGAAATGACATTTTGGAAAACATATTTGATCATCTCC 240
DB GATTGTGATTAAGCATTTTTCAGAAATGACATTTTGGAAAACATATTTGATCATCTCC 959
QY 241 GAAAAAAAACCATTCATTTGTCAGTGTGTGTAAGGGGTTAATTTCTGACAAACCTTG 300
DB GAAAAAAAACCATTCATTTGTCAGTGTGTGTAAGGGGTTAATTTCTGACAAACCTTG 1019
QY 301 AAAAGCATGAATTCACCAATACAAAGTCAATTAATGTAATGTAATGTAATGTAATG 360
DB AAAAGCATGAATTCACCAATACAAAGTCAATTAATGTAATGTAATGTAATGTAATG 1079
QY 361 GCATTTTAAACATCATCTTAAAGACATCATATATTAATCTGTCATGTAAGAAACATTA 420
DB GCATTTTAAACATCATCTTAAAGACATCATATATTAATCTGTCATGTAAGAAACATTA 1139
QY 421 AGGTGTAACATGTAATTAAGTTTCACTGACCTTCAAAATAGCAACAATAATTA 480
DB AGGTGTAACATGTAATTAAGTTTCACTGACCTTCAAAATAGCAACAATAATTA 1140
QY 481 AAAACATCATGTCGATCTCTGCTTATGATGATCATCTCGCTGTTTAAATTTTC 540
DB AAAACATCATGTCGATCTCTGCTTATGATGATCATCTCGCTGTTTAAATTTTC 1259
QY 541 CAAACTTGGTCAGTATTAACATTTTCATTAACAACATGTCATCAAACTTAATATGCTCT 600
DB CAAACTTGGTCAGTATTAACATTTTCATTAACAACATGTCATCAAACTTAATATGCTCT 1260
QY 1260 CAAACTTGGTCAGTATTAACATTTTCATTAACAACATGTCATCAAACTTAATATGCTCT 1319
DB CAAACTTGGTCAGTATTAACATTTTCATTAACAACATGTCATCAAACTTAATATGCTCT 601
QY 601 AAAATGTGTAAGGTTGTGTTGGAAAAAGGTTTATCTTCAATATGTTAAGTCATGAT 660
DB AAAATGTGTAAGGTTGTGTTGGAAAAAGGTTTATCTTCAATATGTTAAGTCATGAT 1320
QY 1320 AAAATGTGTAAGGTTGTGTTGGAAAAAGGTTTATCTTCAATATGTTAAGTCATGAT 1379
DB AAAATGTGTAAGGTTGTGTTGGAAAAAGGTTTATCTTCAATATGTTAAGTCATGAT 661
QY 661 GATTTCATCATGATCAAAATATGAGACTTGTGATTAATGATGATGATGATGATGATGAT 720
DB GATTTCATCATGATCAAAATATGAGACTTGTGATTAATGATGATGATGATGATGATGAT 1380
QY 1380 GATTTCATCATGATCAAAATATGAGACTTGTGATTAATGATGATGATGATGATGATGAT 1439
DB GATTTCATCATGATCAAAATATGAGACTTGTGATTAATGATGATGATGATGATGATGAT 721
QY 721 AAAAATGAATTAAGTGAACATTAATATCTTCCATGATGATGATGATGATGATGATGAT 780
DB AAAAATGAATTAAGTGAACATTAATATCTTCCATGATGATGATGATGATGATGATGAT 1440
QY 1440 AAAAATGAATTAAGTGAACATTAATATCTTCCATGATGATGATGATGATGATGATGAT 1499
DB AAAAATGAATTAAGTGAACATTAATATCTTCCATGATGATGATGATGATGATGATGAT 781
QY 781 TTAAGGAAATCTGAAGTGAAGAAATTAAGAGAACCTTAATGATCAAGATGAAATTAAT 840
DB TTAAGGAAATCTGAAGTGAAGAAATTAAGAGAACCTTAATGATCAAGATGAAATTAAT 1500
QY 1500 TTAAGGAAATCTGAAGTGAAGAAATTAAGAGAACCTTAATGATCAAGATGAAATTAAT 1559
DB TTAAGGAAATCTGAAGTGAAGAAATTAAGAGAACCTTAATGATCAAGATGAAATTAAT 841
QY 841 AATTTCATGATTAAGAAACAGAGAAATTAAGTGAAGAGATGAAGAGATGAAGAA 900
DB AATTTCATGATTAAGAAACAGAGAAATTAAGTGAAGAGATGAAGAGATGAAGAA 1560
QY 1560 AATTTCATGATTAAGAAACAGAGAAATTAAGTGAAGAGATGAAGAGATGAAGAA 1619

QY 901 GATAGCTAGATGAAGAAAGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 960
DB GATAGCTAGATGAAGAAAGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1620
QY 961 AATAATCATCTTACGCTCTTCTTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGA 1020
DB AATAATCATCTTACGCTCTTCTTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGA 1680
QY 1680 AATAATCATCTTACGCTCTTCTTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGA 1739
DB AATAATCATCTTACGCTCTTCTTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGA 1021
QY 1021 AGTGGAGAAAGATCAATTTGCTTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGA 1080
DB AGTGGAGAAAGATCAATTTGCTTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGA 1740
QY 1740 AGTGGAGAAAGATCAATTTGCTTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGA 1799
DB AGTGGAGAAAGATCAATTTGCTTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGA 1081
QY 1081 GATTACGTCGACATTTGGAATGCGATGATGATTAATTAACAAAGATGATGATGATGAT 1140
DB GATTACGTCGACATTTGGAATGCGATGATGATTAATTAACAAAGATGATGATGATGATGAT 1800
QY 1800 GATTACGTCGACATTTGGAATGCGATGATGATTAATTAACAAAGATGATGATGATGAT 1859
DB GATTACGTCGACATTTGGAATGCGATGATGATTAATTAACAAAGATGATGATGATGATGAT 1141
QY 1141 AATAGTATGAAAAAGAAAGAAATCTCAGAGGTGAACCATTTGTTAAAAAGCCAGATG 1200
DB AATAGTATGAAAAAGAAAGAAATCTCAGAGGTGAACCATTTGTTAAAAAGCCAGATG 1860
QY 1860 AATAGTATGAAAAAGAAAGAAATCTCAGAGGTGAACCATTTGTTAAAAAGCCAGATG 1919
DB AATAGTATGAAAAAGAAAGAAATCTCAGAGGTGAACCATTTGTTAAAAAGCCAGATG 1201
QY 1201 GATTATGTCGCAATGAATGAATCATGATGATTTCTGCA 1236
DB GATTATGTCGCAATGAATGAATCATGATGATTTCTGCA 1920
QY 1920 GATTATGTCGCAATGAATGAATCATGATGATTTCTGCA 1955
DB GATTATGTCGCAATGAATGAATCATGATGATTTCTGCA 1955

RESULT 2

ID AB232206
AB232206 strand; DNA, 1239 BP.

XX AB232206;

XX 30-JAN-2003 (first entry)

XX DE Candida albicans essential gene SEQ ID NO 6493.

XX KW Fungus; yeast; tetracyclin; promoter; GRACE strain; biosynthesis;

XX KW signal transduction; DNA replication; cell division; growth;

XX KW proliferation; Candida albicans; fungicide; antifungal; gene; ss.

XX OS Candida albicans.

XX PN WO200253728-A2.

XX PD 11-JUL-2002.

XX PR 26-DEC-2001; 2001MO-US049486.

XX PR 29-DEC-2000; 2000US-0259128P.

XX PR 20-FEB-2001; 2001US-0079202A.

XX PR 22-AUG-2001; 2001US-0314050P.

XX PA (ELITRA) ELITRA PHARM INC.

XX PI Roemer T, Jiang B, Boone C, Bussey H, Ohlsen KL;

XX DR WPI; 2002-56694/60.

XX DR P-PsDB; ABP73656.

XX PT Constructing strains for identifying gene products as effective targets

XX PT for therapeutic intervention, by inactivating in the strain one allele of

XX PT a gene and placing other allele of the gene under conditional expression.

XX PS Claim 37; SEQ ID NO 6493; 167bp + Sequence Listing; English.

XX CC The invention relates to constructing (M1) a strain of diploid fungal

XX CC cells in which both alleles of a gene are modified, comprising modifying

XX CC one allele by insertion or replacement by a cassette having an

XX CC expressible selectable marker and modifying other allele by

XX CC recombination, of a promoter replacement fragment with a heterologous

XX CC promoter, so that expression of the second allele is regulated by the

XX CC promoter. (M1) is useful for constructing a strain of diploid fungal

XX CC cells in which both alleles of a gene are modified. The diploid fungal

CC cells having both alleles modified are useful for identifying a gene that
CC is essential to the survival or growth of a fungus, a gene that
CC contributes to the virulence and/or pathogenicity of a fungus, a gene
CC that contributes to the resistance of a diploid fungus to an antifungal
CC agent, an antifungal agent that inhibits the growth of a diploid fungus
CC and for identifying a therapeutic agent for treatment of a mammalian
CC disease. (M) is useful for identifying a compound which modulates the
CC activity of a gene product, preferably enzymatic activity, carbon
CC compound catabolism, biosynthetic, transporter, transcriptional,
CC translational, signal transduction, DNA replication and cell division
CC activity. The method is useful for identifying a compound having the
CC ability to inhibit growth or proliferation of C. albicans cells and for
CC treating infection by C. albicans. The present sequence is that of an
CC essential Candida albicans gene used in the method of the invention.
CC Note: The sequence data for this patent is not represented in the printed
CC specification but is based on sequence information supplied to Derwent by
CC the European Patent Office

XX Sequence 1239 BP; 481 A; 171 C; 215 G; 372 T; 0 U; 0 Other;

Query Match 99.7%; Score 1232.8; DB 6; Length 1239;
Best Local Similarity 99.8%; Pred. No. 5.5e-233;
Matches 1234; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 ATGAGGAAAGTACGAAACCAATGATCATCTTTAATATCTTCTTCTTCATCA 60
DB 1 ATGAGGAAAGTACGAAACCAATGATCATCTTTAATATCTTCTTCTTCATCA 60
QY 61 CGTCCAAAAGTATATTTGCAATATGAAAGGTGTGATTAAGCCATTAATGACCATCA 120
DB 61 CGTCCAAAAGTATATTTGCAATATGAAAGGTGTGATTAAGCCATTAATGACCATCA 120
QY 121 TTAATTGAGCAAAATTTAAGAACCCACATATGATCGACCGGTATTAATGTAAGAGGAC 180
DB 121 TTAATTGAGCAAAATTTAAGAACCCACATATGATCGACCGGTATTAATGTAAGAGGAG 180
QY 181 GATTGTGATTAAGCATTTTTCAGAAAATCATATTTGGAACACATATGATCATATTC 240
DB 181 GATTGTGATTAAGCATTTTTCAGAAAATCATATTTGGAACACATATGATCATATTC 240
QY 241 GAAAAAACCATTTCATTTGTCAGTGTGTAAGGGGTAAATTCGACAAACATTG 300
DB 241 GAAAAAACCATTTCATTTGTCAGTGTGTAAGGGGTAAATTCGACAAACATTG 300
QY 301 AAAAGCATGAATTCACCCATCAAAAGTCATTTAATGTCATTTGAAAATTTGCAAGAA 360
DB 301 AAAAGCATGAATTCACCCATCAAAAGTCATTTAATGTCATTTGAAAATTTGCAAGAA 360
QY 361 GCATTTTAAATCAATCTTTAAGACATCATATATTTATCTGTCAATGAAAAACATTA 420
DB 361 GCATTTTAAATCAATCTTTAAGACATCATATATTTATCTGTCAATGAAAAACATTA 420
QY 421 ACGTGTAAACATGTATATAAAGTTTTCATCTGACCTTCAAAATAGACAAACATAATTA 480
DB 421 ACGTGTAAACATGTATATAAAGTTTTCATCTGACCTTCAAAATAGACAAACATAATTA 480
QY 481 AAAACATCATGTGTGATCTCTGCTTATCAATGTGATCATCTGCTGTTTAAAAATTTTC 540
DB 481 AAAACATCATGTGTGATCTCTGCTTATCAATGTGATCATCTGCTGTTTAAAAATTTTC 540
QY 541 CAAACTTGTCAGTATTAACAATTTCAATTAACAACATGATCCAAAACCTTAATGCTCC 600
DB 541 CAAACTTGTCAGTATTAACAATTTCAATTAACAACATGATCCAAAACCTTAATGCTCC 600
QY 601 AAAATGTGTAAAGTGTGTGGAAGAAAAGTTTATCTTCACATATGTTAAGTCATGAT 660
DB 601 AAAATGTGTAAAGTGTGTGGAAGAAAAGTTTATCTTCACATATGTTAAGTCATGAT 660
QY 661 GATTCTACCATGATCAAAAATATGAGCTGTGATTTATGTGATGTGGGAAAATTTGCAAG 720
DB 661 GATTCTACCATGATCAAAAATATGAGCTGTGATTTATGTGATGTGGGAAAATTTGCAAG 720
QY 721 AAAAATGAATTAAGTGAACATTAATAATCTTCATGATGTGAATAATCCCTGATGATTA 780

DB 721 AAAAATGAATTAAGTGAACATTAATAATCTTCATGATGTGAATAATCCCTGATGATTA 780
QY 781 TTAAGGAACATCGAATGAAAAAATTAAGAACCTTTGATCAAGAGATCGAAATTAAT 840
DB 781 TTAAGGAACATCGAATGAAAAAATTAAGAACCTTTGATCAAGAGATCGAAATTAAT 840
QY 841 AATTGTCATGAATTTAGAAACAGAAATTAAGTGAAGAGATGAAGATGAAGAA 900
DB 841 AATTGTCATGAATTTAGAAACAGAAATTAAGTGAAGAGATGAAGATGAAGAA 900
QY 901 GATAGCTATGATGAAAAAAGATGATGATGATGATGATGATGATGATGATGATGATGAT 960
DB 901 GATAGCTATGATGAAAAAAGATGATGATGATGATGATGATGATGATGATGATGATGAT 960
QY 961 ATTAATATCTTTTACTCTCTTTTGGAAAGGTTCAGAAAGTGTTCATTAATCTTGAAT 1020
DB 961 ATTAATATCTTTTACTCTCTTTTGGAAAGGTTCAGAAAGTGTTCATTAATCTTGAAT 1020
QY 1021 AGTGGGAGAGATGCAATTTGCTCTAAGATTAATGATGATGATGATGATGATGATGAT 1080
DB 1021 AGTGGGAGAGATGCAATTTGCTCTAAGATTAATGATGATGATGATGATGATGATGAT 1080
QY 1081 GATTTCATGTCGACATTTGAAATGCGATGATGATGATGATGATGATGATGATGATGAT 1140
DB 1081 GATTTCATGTCGACATTTGAAATGCGATGATGATGATGATGATGATGATGATGATGAT 1140
QY 1141 AATAGTATGAAAAAGAAAGAACTCCAGAAAGTGAACATTTGTTAAAAAGCCAGATG 1200
DB 1141 AATAGTATGAAAAAGAAAGAACTCCAGAAAGTGAACATTTGTTAAAAAGCCAGATG 1200
QY 1201 GATTATTTGCCAAATGAACATCAGTATTTCTCGA 1236
DB 1201 GATTATTTGCCAAATGAACATCAGTATTTCTCGA 1236

RESULT 3
ADA53158
ID ADA53158 standard; cDNA; 2132 BP.
XX
AC ADA53158;
XX
DT 20-NOV-2003 (first entry)
XX
DE Human coding sequence, SEQ ID 726.
XX
XX Cytostatic; Anti-inflammatory; Osteopathic; Neuroprotective; Nocotropic;
KW Gene Therapy; human; secretory protein; membrane proteins; cancer;
KW Inflammatory disease; osteoporosis; neurological disease; gene; ss.
XX
OS Homo sapiens.
XX
XX
PN EF1293569-A2.
XX
PD 19-MAR-2003.
XX
PF 21-MAR-2002; 2002EP-00006586.
XX
PR 14-SEP-2001; 2001JP-00328381.
PR 24-JAN-2002; 2002US-0350435P.
PA
PA (HELI-) HELIX RES INST.
PA (REAS-) RES ASSOC BIOTECHNOLOGY.
XX
PI Isogai T, Sugiyama T, Otsuki T, Wakamatsu A, Sato H, Ishii S,
PI Yamamoto J, Isono Y, Hio Y, Otsuka K, Irie R, Tamechika I,
PI Seki N, Yoshikawa T, Otsuka M, Nagahari K, Masuno Y;
XX
DR WPI; 2003-395539/38.
DR P-PDB; ADA54797.
XX
PT New polynucleotides encoding full-length polypeptides, e.g. secretory
and/or membrane proteins, useful for developing medicines for diseases in

PT which the gene is involved, or as target molecules for gene therapy.
XX
PS Claim 1; SEQ ID NO 726; 205bp; English.
XX
CC The present invention relates to novel human secretory or membrane
CC protein (ADA54072-ADA55710) and their coding sequences (ADA52433-
CC ADA54071). The coding sequences are useful in the gene therapy of
CC diseases caused by abnormalities of the proteins, e.g. cancer,
CC inflammatory diseases, osteoporosis or neurological disease.
XX
SQ Sequence 2132 BP; 833 A; 423 C; 347 G; 529 T; 0 U; 0 Other;

Query Match 8.0%; Score 99.4; DB 10; Length 2132;
Best Local Similarity 50.6%; Pred. No. 4.7e-10;
Matches 297; Conservative 0; Mismatches 281; Indels 9; Gaps 2;

QY 7 GAAAGTGAAGAAACCAATCGATATCATCTTAAATCTCTCTCTTCATCAGCTGCC 66
DB 1384 GAATGTGGAAAGCTTTTAAACCAATCTCAATCTTACTACATAGAAATTCATCT 1443
QY 67 AAAAGTATATTTGACATATGAGGGTGTGATTAAGCCTATTAATGACCATCATTTA 126
DB 1444 GAGAGAAATCTTACAAATGTGAGAAATGGCAAGCTTTCTATGATCTCAAACT 1503
QY 127 GAGCAATTTAAGAAACCAAGTAATGATCGACCTATTAATGTACAGTGAAGTGT 186
DB 1504 ACTGAACATAGAAATTTATCTGAGAGAAACCTTACACATGT-----GAGAAATGT 1557
QY 187 GATTAAGCATTTTTCAGAAATTCATTTGGAAACATATTTGATACATTTCCGAAAA 246
DB 1558 GCGAAAGCTTTTAAACCACTCTTCACTTGTACATTAAGGTAAATTAATCTGAGAG 1617
QY 247 AAACCATTCATTTGTCAGTGTGTGTAAGGGGTAAATTTCTGACAACTTGAAGAA 306
DB 1618 AAACCTTACCAATGTGAAGAAATGTGTAAAGCTTTAAACAGTCTTCACTTACTA 1677
QY 307 CATGAATACCCATACAAAGTCATTTAAATGTACATTTGAAATTTGCAAGAGCATTT 366
DB 1678 CATTAAGAAATTCATCTGAGAGAAACCTTACCAATGTGAAATTAATGTGCAAGCTTTT 1737
QY 367 TTTAAACATCATCTTTAA---GACATCATATATTTATCTGTCATGAAATTAACATAAG 423
DB 1738 AACCGTCTCTAAACCTTACTGACATTAAGAAATTCATCTGTGTAGAAACCTTACAA 1797
QY 424 TGTAAACAATGTAAATTAAGTTTCACTCACCTTCAAAATTTAGCAACATTAATTA 483
DB 1798 CCTAAAGATGTAAAGTATTTTGAAGACCTTCAAAATTTCTTAAACATTAAGAAAT 1857
QY 484 CATCATGTGTGATCTCTCTTATCATGTATCATCTGTGTGTTTAAATTTTCAA 543
DB 1858 TATGCTGTGTGAGAAATCTTGAAGAAATGTGAAGATGTAAACCTTTTAAAGTGTGAC 1917
QY 544 ACTTGTCAGTATTAATTTCAATTAATAAACAATGATCCAAACT 590
DB 1918 ACTTGATGTGTGATTAATTTCAATTAATAAATAAATAAATCTTAAGT 1964

RESULT 4
ACA98970
ID ACA98970 standard; cDNA, 1705 BP.
XX
XX
AC ACA98970;
XX
DT 25-JUL-2003 (first entry)
XX
DE cDNA encoding human nucleic acid-associated protein (NAAP) #51.
XX
KW Human; nucleic acid-associated protein; cytosolic; antiarteriosclerotic;
KW anticonvulsant; nootropic; neuroprotective; cerebroprotective; anti-HIV;
KW antiallergic; antiinflammatory; thyromimetic; gene therapy;
KW cell proliferative disorder; cancer; atherosclerosis;
KW neurological disorder; epilepsy; Huntington's disease; stroke;
KW immune disorder; inflammatory disorder; AIDS; allergy;

KW developmental disorder; Hypothyroidism; Cushing's syndrome; infection;
KW protein-protein interaction; drug-target interaction;
KW gene expression profile; gene; ss.
XX
OS Homo sapiens.
XX
XX WO2003023003-A2.
XX
XX
PD 20-MAR-2003.
XX
XX
PF 05-SEP-2002; 2002MO-US028540.
XX
XX
PR 07-SEP-2001; 2001US-0317792P.
XX
PR 07-SEP-2001; 2001US-0317912P.
XX
PR 14-SEP-2001; 2001US-0322270P.
XX
PR 21-SEP-2001; 2001US-0324040P.
XX
PR 28-SEP-2001; 2001US-0326732P.
XX
PR 19-OCT-2001; 2001US-0346715P.
XX
PR 25-JAN-2002; 2002US-0351749P.
XX
PR 22-FEB-2002; 2002US-0359498P.
XX
XX
PA (INCY-) INCYTE GENOMICS INC.
XX
XX
PI Tang YT, Jackson JL, Griffin JA, Elliott VS, Forsythe IJ;
PI Becha SD, Richardson TW, Lee EA, Sprague WW, Emerling BW;
PI Thangavelu K, Warren BA, Tran UK, Yue H, Xu Y, Yue H, Li JX;
PI Hafalia AD, Sanjwal B, Margulis JP, Gortvad AE, Lee SY, Ison CH;
PI Baugh MR, Chawla NK, Nguyen DB, Swannaker A, Zebairadian Y, Shah P;
PI Thornton M, Yao MG, Khan FA, Gandhi AR, Yang J, Kabie AE;
PI Burford N, Ramkumar J;
XX
XX
DR WPI; 2003-313243/30.
XX
XX P-PsDB; AB096722.
XX
XX
PT New human nucleic acid associated proteins (NAAP), useful for diagnosing,
PT creating and preventing diseases or conditions associated with the
PT aberrant NAAP expression e.g. cancer, AIDS, atherosclerosis, epilepsy, or
PT infections.
XX
XX
PS Claim 5; Page 340-341; 345pp; English.
XX
XX
CC The invention describes a novel human isolated nucleic acid-associated
CC polypeptide (NAAP). The polypeptides and polynucleotides are useful in
CC diagnosing, treating and preventing diseases or conditions associated
CC with the decreased expression or overexpression of NAAP, such as cell
CC proliferative (e.g. cancer, atherosclerosis), neurological (e.g.
CC epilepsy, Huntington's disease, stroke), immune/inflammatory (e.g. AIDS,
CC allergies) and developmental (e.g. Hypothyroidism, Cushing's syndrome)
CC disorders, or infections. These are also useful in assessing the effects
CC of exogenous compounds on the expression of nucleic acid and amino acid
CC sequences of NAAP. The NAAP or its fragments are useful in screening
CC compounds for effectiveness as agonist or antagonist of the polypeptides,
CC or in altering the expression of the target polynucleotide and compounds
CC that specifically bind to or modulate the activity of the polypeptide.
CC The microarray is useful in monitoring or measuring protein-protein
CC interactions, drug-target interactions, and gene expression profiles.
CC This sequence encodes a novel human nucleic acid-associated protein
CC (NAAP)
XX
XX
SQ Sequence 1705 BP; 622 A; 305 C; 314 G; 464 T; 0 U; 0 Other;

Query Match 7.8%; Score 96.8; DB 9; Length 1705;
Best Local Similarity 52.9%; Pred. No. 1.5e-09;
Matches 259; Conservative 0; Mismatches 222; Indels 9; Gaps 2;

QY 7 GAAAGTGAAGAAACCAATCGATATCATCTTAAATCTCTCTCTTCATCAGCTGCC 66
DB 695 GAATGTCCCAAAACATTTTGAATGCTTTCACACCTCAACTCAACATTAAGATCCAAACT 754
QY 67 AAAAGTATATTTGACATATGAGGGTGTGATTAAGCCTATTAATGACCATCATTTA 126
DB 755 AGAGTGAATTTCTCAAAATGTGAAGACATATGAGAGCCTTTAACTGTGCTCAACCTT 814

QY 127 GAGCAACATTTAAGAACCCAGTAATGATCGACCGTATTAATGTACAGTGAAGATGT 186
 DB 815 AATTAACATTAAGAGATTCTACTGAGAAACCTTACAAATGTAA-----GATGT 868
 QY 187 GATTAAGCATTTTTCAGAAATTCATTGGAAACATATTGTATCATCTTCGAAAA 246
 DB 869 GGCMAACCTTTTACAGACCTCACACCTTATTAAGACATTAAGAAATTCATCTGAAGAG 928
 QY 247 AAACATTCATGTTGATGTTGTTGTTAAAGGGTTAATCTCGACAAACCTTGAAGA 306
 DB 929 AAACCTTAAGAAATGTAAGAAATGTCAGAAAGCTTTTAAACAGTATGACCTTACTACA 988
 QY 307 CATGAATCAACCATCAAGTCAATTTAAATGATCATTTGAAATTTGCAAGAACATTT 366
 DB 989 CATTAATTAATTAATTAATGAGGAAATTCCTTACAAATGTGAGAAATGTGTAAGCTTTT 1048
 QY 367 TATTAACATCAATCTTTAAGACATCATATATATCTGTT---CATGAAAAACATTAACG 423
 DB 1049 AACCAACCTCAAGTCTACTGAAACATTAATTAATCATCCGAGAGAAATGCTACAA 1108
 QY 424 TGTAAACATGTAATTAAGTTTTCATCGACCTTCAAAATTAAGACAACTTAATTAA 483
 DB 1109 TGTGAAGAAATGTGGCAAAAGTTTAAACCGATCTTCAAACTTACAGACATTAATAATTT 1168
 QY 484 CATCATGTG 493
 DB 1169 CATTAAGAG 1178

RESULT 5
 ACAS6456
 ID ACAS6456 standard; cDNA; 2320 BP.
 XX

ACAS6456;
 06-JUN-2003 (first entry)

Human signalling pathway polynucleotide probe SEQ ID NO 1054.

Human; probe; ss; array element; Parkinson's disease;
 signalling pathway population; cancer; adenocarcinoma; leukaemia;
 immunopathy; AIDS; asthma; neuropathy; Alzheimer's disease; microarray.

Homo sapiens.

US6500938-B1.

31-DEC-2002.

30-JAN-1998; 98US-00016434.

30-JAN-1998; 98US-00016434.

(INCYTE GENOMICS INC.

Au-Young J, Sellhammer JJ;

WPI; 2003-352189/33.

Combination of polynucleotide probes, useful as array elements in a
 microarray for monitoring the expression of a number of target
 polynucleotides.

Claim 1, SEQ ID NO 1054; 65pp; English.

The invention relates to a combination which, comprises a number of
 polynucleotide probes comprising a sequence selected from one of the 1490
 sequences mentioned in the specification. The combination is useful as an
 array element in a microarray for monitoring the expression of a number
 of target polynucleotides. The microarray is particularly useful in the
 diagnosis and treatment of cancer and immunopathology and neuropathology.
 The microarray is useful in diagnostics and treatment regimens, drug
 discovery and development, toxicological and carcinogenicity studies,

CC forensics and pharmacogenomics. The microarray is also useful for
 CC monitoring progression of diseases and for developing sophisticated
 CC profiles for the effects of currently available therapeutic drugs. The
 CC combination is also useful for purifying a subpopulation of mRNA, cDNAs
 CC and genomic fragments and in research and diagnostic applications. The
 CC array can detect changes in expression in a large number of genes coding
 CC for different signalling pathway populations which can be used to diagnose
 CC various diseases including cancer e.g. adenocarcinoma and leukaemia,
 CC immunopathies e.g. AIDS and asthma, neuropathies e.g. Alzheimer's disease
 CC and Parkinson's disease. The present sequence represents a polynucleotide
 CC probe of the invention. Note: The sequence data for this patent did not
 CC form part of the printed specification but was obtained in electronic
 CC seqdata.uspto.gov/sequence.html?docID=06500938B1
 XX

Sequence 2320 BP; 903 A; 423 C; 412 G; 582 T; 0 U; 0 Other;
 Query Match 7.8%; Score 96.6; DB 10; Length 2320;
 Best Local Similarity 51.3%; Pred. No. 1.7e-09;
 Matches 306; Conservative 0; Mismatches 279; Indels 12; Gaps 3;

QY 69 AAGTATATTTGACATATGAAGGTGTGATTAAGCTATATTCGACATCATATTATGA 128
 DB 964 AGAGAAACCTTACAAATGTAAAGATGTGTTAAAGCTTTTAAACGATCTTCAACCTTAC 1023
 QY 129 GCAACATTTAAGAACCCAGTATGATGACCGTATTAATGATGAGTGA 188
 DB 1024 TACCATTAAGAAATTTCTATCTGAGAGAAACCTTACAAATGT-----GAAGATGTGG 1077
 QY 189 TAAAGCATTTTTCAGAAATTCATTTGGAACATATTTGATCATATTCCGAAAAAA 248
 DB 1078 CAAGCCTTTAAGCAGTCCCAACCTTACTACATCAATTAATTCATCTGAGAGAA 1137
 QY 249 ACCATTCATTTGTCAGTGTGTGTAAAGGGTTAATTCGACAAACATTTGAAAGACA 308
 DB 1138 ACCCTCAATGTAAAGATGTGAAAGAGCTTTTAAACAGTCTGACACCTTACACACA 1197
 QY 309 TGAATCAACCCATCAAGATCATTTAAATGATTAATTTGCAAGACATTTTA 368
 DB 1198 TGAAGTATTCATATCTGAGAGAAACCTTACAAATGTGAAAGATGGAAGCTTTAA 1257
 QY 369 TAAACA---TCAATCTTTAAGACATATATATCTGTTTCATGAAAAAATTTAAACGTG 425
 DB 1258 TCATTTCTGACACCTTACATACATTAAGATTAATTCATCTGAGAGAAACCTTAAACATG 1317
 QY 426 TAAACATGTAATTAAGTTTTCATGACCTTCAAAATTAAGCAACATTAATTAACA 485
 DB 1318 TAAAGATGTGTAAGCTTTTAAACATCTTCAACCTTACATTAAGATTAATTC 1377
 QY 486 TCATGTGATCTCCCTGCTTATCAATGTATCATCTCGTGTGTTTAAAAATTTCCAAAC 545
 DB 1378 TACTGAGAGAGACCTTACAAATCTAAAGATGTGAAAAAGCTTTTAAACCAATCTCAAA 1437
 QY 546 TTGTCAGTATTAATTAATTTCAATTAACACATGTCATCCAAATTAATGTCCTTAATG 605
 DB 1438 ACTTACTGAACATTAAGAAATTCATCTGAGAGAAACCTTA---TGAATGTAAAAAATG 1494
 QY 606 TGGTAAAGTTGTGTGGGAAAAAAGGTTTATCTTCAATATGTTAATGATGATGA 662
 DB 1495 TGGCAAAAGCTTTTAAACAGTCTCAATCTTACTAGACATTAAGAAAGTATACAG 1551

RESULT 6
 ADIS6252
 ID ADIS6252 standard; DNA; 2320 BP.
 XX
 AC ADIS6252;
 XX
 DT 22-APR-2004 (first entry)
 XX
 DE Human polynucleotide probe #1054.
 XX
 KM Human; probe; ss; receptor-like polypeptide; transducing polypeptide;

XX	effectore-like polypeptide; cancer; immunopathology; neuropathology;
KW	drug development; toxicology; carcinogenicity; AIDS;
KW	signalling pathway polypeptide; adrenal gland; bladder; bone;
KW	bone marrow; brain; breast; cervix; tumour; immunopathology;
KW	diabetes; pancreatitis; osteoporosis; ulcerative colitis; neuropathology;
XX	dementia; amnesia; epilepsy; Alzheimer's disease; depression.
OS	Homo sapiens.
PN	US2004010136-A1.
XX	
PD	15-JAN-2004.
XX	
PF	26-NOV-2002; 2002US-00305720.
XX	
PR	30-JAN-1998; 98US-00016434.
XX	
PA	(INCY-) INCYTE GENOMICS INC.
XX	
PI	Au-Young J, Sellhammer JI;
XX	
DR	WPI; 2004-090520/09.
XX	
PT	New composition comprising polynucleotide probes, useful as array
PT	elements in a microarray for monitoring the expression of target
PT	polynucleotides or purifying a subpopulation of mRNAs, cDNA, or genomic
PT	fragments.
XX	
XX	Claim 6; SEQ ID NO 1054; 73pp; English.
XX	

	Query Match	7.8%	Score 96.6	DB 12	Length 2320
	Best Local Similarity	51.3%	Pred. No. 1.7e-09		
	Matches 306	Conservative	0	Mismatches 279	Indels 12
				Gaps 3	
Qy	69	AAAGTATTTTGCACATATGAAAGGGTGTGATAAAGCCCTATATCGACCATCATTTATTTGA	128		
Db	964	AGAGAAACCCTACAAATGATTAAGATGTGTAAAGCTTTTAAACGATCTTCAACCCCTTAC	1023		
Qy	129	GCAACATTTAAGAACCCACAGTAATGATGACCGTATAAATGTACAGTGCAGATTGTGA	188		
Db	1024	TACCCATGAAAAAATTCATCTGAGAGAAACCCTTACAAATGT-----GAAAGATGTGG	1077		
Qy	189	TAAAGCATTTTTCAGAAAAATCACATTTTGAAACACATATTGTATACATTTCCGAAAAAAA	248		
Db	1078	CAAAAGCCTTTTAGCAGTCCTTAAACCTTACTACACATTAAGATTAATTCATCTGAGAGAA	1137		

Qy	249	ACCAATTCATTGTTCACTGATGTGGTAAAGGGGTTAAATTCGACAAACCTTGAAGACA	308
Db	1138	ACCCTTCAAAATGTAAAAAATGTGAAAGAGCCCTTTAACCAAGTCTGCACACCTTACACACA	1197
Qy	309	TGAATATCACCATACAAAGTCATTTAAATGTACATTTGAAAATTTGTCAAGAGCATTTTA	368
Db	1198	TGAGGTAAATTCATCTGAGAGAAACCCCTCAATATGTGAAAAAATGTGAAAGAGCTTTAA	1257
Qy	369	TAAACA---TCAATCTTTAAGACATCAATATTAATCTGTTCATGAAAAAACATTAACGTG	425
Db	1258	TCATTTCTCACACCTTACTACACATTAAGATTAATTCATACGTGAGAAAGACCTTCAAAATG	1317
Qy	426	TAAACATGTAAATTAAGATTTTCACTGCACCTTCAAAATTTAGACAACATAAATTAAGACA	485
Db	1318	TAAAGATATGTAAAGCTTTTAAACCTCTTCAACCTTACTTAACACTTAAGATTAATTTCA	1377
Qy	486	TCAATGTGTGATCTCTCGCTTATCAATGTGATTCATCTCGTGTGTTTAAAAATTTCCAAC	545
Db	1378	TACTGAGAGAAAGCTTACAAATCTAAAGATGTGAAAAAGCTTTTAAACCATCTCBA	1437
Qy	546	TTGGTCAGTATTCAAATTCATATTAACAAACATGCAATCCAAAACCTTAATATGCTCTAATG	605
Db	1438	ACTTACTGAACATATGAAGAAATTCATACTGAGAGAAACCTTA---TGAATGTGAAAAATG	1494
Qy	606	TGCTAAAGGTGTGTGGGAAAAAGTTTATCTTCAATATGTTAAGTACATGATGA	662
Db	1495	TGGGAAAGCTTTTAAACAGTCTCCAAATCTTACTAGCATTAAGAAAAAGTCAATACAGA	1551

CC	RESULT 7
XX	AAS91317
XX	AAS91317 standard; cDNA; 4563 BP.
AC	AAS91317;
XX	
DT	13-FEB-2002 (first entry)
XX	
DE	DNA encoding novel human diagnostic protein #27121.
XX	
KW	Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW	food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX	
OS	Homo sapiens.
XX	
PN	WO200175067-A2.
PD	11-OCT-2001.
XX	
PF	30-MAR-2001; 2001WO-US008631.
XX	
XX	31-MAR-2000; 2000US-00540217.
PR	23-AUG-2000; 2000US-00649167.
XX	
PA	(HYSE-) HYSEQ INC.
XX	
PI	Drmamac RT; Liu C, Tang YT;
XX	
DR	WPI; 2001-639362/73.
DR	P-PSDB; ABG271130.
XX	
PT	New isolated polynucleotide and encoded polypeptides, useful in
PT	diagnostics, forensics, gene mapping, identification of mutations
PT	responsible for genetic disorders or other traits and to assess
PT	biodiversity.
XX	
PS	Claim 1; SEQ ID NO 27121; 103bp; English.
XX	
XX	
CC	The invention relates to isolated polynucleotide (I) and polypeptide (II)
CC	sequences. (I) is useful as hybridisation probes, polymerase chain
CC	reaction (PCR) primers, oligomers, and for chromosome and gene mapping,
CC	and in recombinant production of (II). The polynucleotides are also used
CC	in diagnostics as expressed sequence tags for identifying expressed
CC	genes. (I) is useful in gene therapy techniques to restore normal

CC activity of (II) or to treat disease states involving (II). (II) is
CC useful for generating antibodies against it, detecting or quantitating a
CC polypeptide in tissue, as molecular weight markers and as a food
CC supplement. (II) and its binding partners are useful in medical imaging
CC of sites expressing (II). (I) and (II) are useful for treating disorders
CC involving aberrant protein expression or biological activity. The
CC polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. A564197-A564564 represent novel human diagnostic
CC coding sequences of the invention. Note: The sequence data for this
CC patent did not appear in the printed specification, but was obtained in
CC electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences

XX Sequence 4563 BP; 1752 A; 814 C; 824 G; 1173 T; 0 U; 0 Other;

XX Query Match 7.7%; Score 95.2; DB 5; Length 4563;

XX Best Local Similarity 52.7%; Pred. No. 3.5e-09;

XX Matches 258; Conservative 0; Mismatches 223; Indels 9; Gaps 2;

QY 7 GAAGTGCAGAAACCAATGATATCATCTTAAATCTTCTTTCATCAGCCTCC 66

Db 1315 GAATGGGCAAGCCTTTACAGCTCCTCAACCTTACTCAACATTAACATTCATCT 1374

QY 67 AAAAAGTATATTTGCATATGAAAGGTGTGATTAAGCTTATATGACCATCATTA 126

Db 1375 GGAGGCAAAACCTACAAATGTGAAGATGTGGCAAGCTTTTAAACATCTTTCAGCCCTT 1434

QY 127 GACCAACATTTAAGAACCCACAGTAATGATGACCGTAAATGATGACGAGCATTTG 186

Db 1435 ACTTAACATTAAGTATATTTCTACTGCGGAGAAACCATACAAATGT-----GAAGATGT 1488

QY 187 GATTAAGCATTTTTCAGAAATCACAATTGAAACATATTTGATACATTTCCGAAAA 246

Db 1489 GCGAAGCTTTTGGCAATCTCTACACCTTACTAGCATTAAGCAATTCATCTGAGAG 1548

QY 247 AAACCATTTCCATTTGTCAGTGTGTGTGAAGGGGTTAATTTCTGACAAACCTTGA 306

Db 1549 AAACCCCTACAAATGTGAAGATGTGGCAAGCTTTTAAACATTTCTCAGACCTTAGA 1608

QY 307 CATGAATACCAATCAAGTCAATTTAATGTACATTTGAATTTGTCAAGAGCATTT 366

Db 1609 CATTAAGTATATTTACTGAGAAAGAAACCTTACAAATGTGAAGATGTGGAAAGCTTTT 1668

QY 367 TATTAACATCAATCTTTAAGA---CATCATATATTTCTGTTCATGAAAAAACATTA 423

Db 1669 AGCGAGCTTCACACCTTAAAGCAACATGATATTCATCTGAGAGAAACCTTACAA 1728

QY 424 TGTAAACATGTATTAAGTTTTCATCTGACCTTCAAAATTTAGACAAACATTAATT 483

Db 1729 TGTGAAGATGTGTGAAGCTTTTAAGTGTTCATCAAAACCTTACTGATCAATTA 1788

QY 484 CATCAGGTG 493

Db 1789 CATACTGGAG 1798

RESULT 8

ID ADN04717 standard; cDNA; 981 BP.

XX ADN04717;

DT 01-JUL-2004 (first entry)

DS Antipsoriatic cDNA sequence #572.

XX ds; gene; antipsoriatic; gene therapy; psoriasis; diagnosis.

XX Homo sapiens.

PN WO2004028479-A2.

XX 08-APR-2004.

XX 25-SEP-2003; 2003WO-US030907.

XX 25-SEP-2002; 2002US-0414006P.

XX (GENTH) GENENTECH INC.

XX Bodary S, Clark H, Jackman J, Schoenfeld J, Williams PM, Wood WJ,

XX Wu JD;

XX WPI; 2004-305105/28.

XX P-PSDB; ADN04718.

XX New PRO nucleic acid or polypeptide, useful for preparing a

XX pharmaceutical composition for diagnosing or treating psoriasis in a

XX mammal.

XX Claim 1; SEQ ID NO 1111; 3069bp; English.

XX The invention relates to novel polynucleotide and polypeptides for

XX treating psoriasis or a sequence having at least 80% identity to the

XX above sequences. The nucleic acid is useful for preparing a composition

XX for diagnosing or treating psoriasis in a mammal. This sequence

XX corresponds to one of the polynucleotides of the invention.

XX Sequence 981 BP; 380 A; 178 C; 170 G; 253 T; 0 U; 0 Other;

XX Query Match 7.7%; Score 95; DB 12; Length 981;

XX Best Local Similarity 50.4%; Pred. No. 3.1e-09;

XX Matches 289; Conservative 0; Mismatches 275; Indels 9; Gaps 2;

QY 89 AAGGCTGTATTAAGCCTTAATTCAGACATCATTTATTAGACAACTTAAGAACCCACA 148

Db 1 AAGAAATGTGCAAGAGCTTTAGATATTTCTCAACCTTACTTAACATTAAGATTAATTCAT 60

QY 149 GTATATGATGACCGGTATTAATGATGACGATGATGATGATGATGATGATGATGAT 208

Db 61 CTGGAAGAAACCTTCAAAATGCA-----ATGAATGTGTAAAGCTTTAACTGTCTCT 114

QY 209 CACATTTGAAACACATATTTGATATCATTCGAAAAAAACCATTCATGTTCAAGT 268

Db 115 CAACCTTTACTTAACATTAAGATTCATCTGAGAGAAACCTTCAAAATGTGAAGAT 174

QY 269 GTGTAAAGGGGTTAATTTCTGACCAACCTTGAAGAGATGAATCAACCATTAACAAGT 328

Db 175 GTGGCAAGCTTTTAAACAGTCTCAACCTTACTGACATTAAGATTAAGTCAATCTGAG 234

QY 329 CATTTAATGTACATTTGAAATTTGCAAGAGCATTTTATTAACATCAATCTTTAAGC 388

Db 235 AGAAACCTTCAAAATGTGAAGATGTGTAAAGCTTTAAACGATCAACCTTAAC 294

QY 389 ATCATATATTTATCTGTCATGAAGAAACATTAACGTTAAACATTAATTAAGTTTCA 448

Db 295 ATTAAGAAATTTATTAAGAAACCAATCAATGTGAAGATGTGAAGAAAGCTTTA 354

QY 449 CTCGACCTTCAAAATTAAGCAACATTAATTAACATCATGATGATCTCTGCTTATC 508

Db 355 GTGTATTTCAACCTTACTTAAGCAATTAAGTATTTACTGAGCAAAACCTTACAAAT 414

QY 509 AATGTATCATCTCTGTTGTTTAAATTTCCAAACCTTGGTCAATTAATTTCA 568

Db 415 GTGACGAATGTGCAAGTGTCTTGAAGCAATTCACACCTTACTGAACTTAAGAGTTC 474

QY 569 TAAACACATCTGCACTCAAAACCTTAATGTCTTAATGTGTAAAGTGTGGGAAA 628

Db 475 ATACTGAGAGAAACCTTAC---AAATGCAATGAATGTGTAAAGCTTTAACTGATCT 531

QY 629 AAGTTTATCTTCATATGATTTAAGTCATGATG 661

Db 532 CAACCTTACTTAACATTAAGAGATTCATCTG 564

RESULT 9
AAL51569
ID AAL51569 standard; DNA; 2597 BP.
XX
AC AAL51569;
XX
DT 10-APR-2003 (first entry)
XX
DE Human nucleic acid-associated protein coding sequence - SEQ ID NO 52.
XX
KW Human; gene; db; nucleic acid-associated protein; NAAP; arteriosclerosis;
cell proliferative disorder; atherosclerosis; cirrhosis; hepatitis; AIDS;
cancer; developmental disorder; renal tubular acidosis; anemia; asthma;
mental retardation; neurological disorder; Alzheimer's disease; epilepsy;
Parkinson's disease; autoimmune disorder; inflammatory disorder; allergy;
Crohn's disease; transgenic animal; animal model.
XX
OS Homo sapiens.
XX
PN WO200300864-A2.
XX
PD 03-JAN-2003.
XX
PF 20-JUN-2002; 2002WO-US021179.
XX
PR 22-JUN-2001; 2001US-0300518P.
PR 29-JUN-2001; 2001US-0301787P.
PR 29-JUN-2001; 2001US-0301792P.
PR 29-JUN-2001; 2001US-0301892P.
PR 29-JUN-2001; 2001US-0301893P.
PR 06-JUL-2001; 2001US-0303405P.
PR 06-JUL-2001; 2001US-0303442P.
PR 15-MAR-2002; 2002US-0364438P.
XX
PA (INCYTE GENOMICS INC.
XX
PI Gandhi AR, Swarnakar A, Hafalia AJA, Warren BA, Emerling BM;
Arvizu CS, Ison CH, Honchell CD, Lee EA, Yue H, Forsythe IU;
Rakumar J, Griffin JA, Yang J, Sanjanwala NM, Baughn MR;
Borowsky ML, Yao MG, Wallia NK, Bandman O, Lai PG, Becha SD, Lee SY;
Richardson TW, Elliott VS, Luo W, Tang YT, Zebajadian Y, Lu Y;
XX
DR WPI; 2003-201420/19.
DR P-PSDB; AAO16419.
XX
PT New nucleic acid-associated proteins and polynucleotides, useful for
diagnosing, treating or preventing cell proliferative (e.g. cancer),
neurological (e.g. epilepsy or Parkinson's disease), or autoimmune
disorders (e.g. AIDS).
PT
PS Claim 12; Page 292-293; 312pp; English.
XX
XX The invention comprises the amino acid and coding sequences of human
nucleic acid-associated proteins (NAAP). The DNA and protein sequences of
the invention are useful for diagnosing, treating or preventing disorders
associated with aberrant expression of NAAP, such as: cell proliferative
disorders (e.g. arteriosclerosis, atherosclerosis, cirrhosis, hepatitis
or cancer); developmental disorders (e.g. renal tubular acidosis, anemia
or mental retardation); neurological disorders (e.g. Alzheimer's disease,
Parkinson's disease or epilepsy); and autoimmune/inflammatory disorders
(e.g. AIDS, allergies, asthma or Crohn's disease). The DNA sequences of
the invention are useful for creating transgenic animals to model human
disease. The present DNA sequence encodes a human nucleic acid-associated
protein of the invention
XX
SQ Sequence 2597 BP; 947 A; 504 C; 505 G; 641 T; 0 U; 0 Other;

Query Match 7.7%; Score 94.6; DB 8; Length 2597;
Best Local Similarity 55.1%; Pred. No. 4.2e-09;
Matches 231; Conservative 0; Mismatches 179; Indels 9; Gaps 2;

QY 78 TTGACATATGAGGGGTGTGATTAAGCCTTAATATGACCATTATTAAGCAATTT 137
DB 1070 TTATTAATGTGAAGATGTGGCAAGCCTTAATGTCTCAACCTTACTCAACATA 1129
QY 138 AAGAACCCACAGTATGATCGACCGGTATTAATATGATGACGATTTGATTAAGCAT 197
DB 1130 GAGAAATTCACTACGAGAGAAACCTTACAAATGT-----GAAAGGTGGCAAAAGCCTT 1183
QY 198 TTTCAAAAAATCACAATTGGAAAACACATATTGTATCATCTCCGAAAAAAACCATTCGA 257
DB 1184 TTAAGTGTCTTCACTTCACTTCAACATTAAGGAATTCATCTGAGAGAAAAACATACAA 1243
QY 258 TTGTTCAAGTGTGTGTAAGGGGTAAATTCCTGACACACTTGAAGAAGCATGAATCAC 317
DB 1244 AAGTAAGAAGATGTGCCAAAGCCTTAACACATCTCCACCTCACACACATATAAAGAT 1303
QY 318 CCATCAAAAGTCATTTAAATGATACATTGAAATTTGTCAAGAGCATTTTATAA--ACA 374
DB 1304 TCATACCGGAGAGAAACCCCTACAAATGTGAAGAAATGTGCAGAAAGCCTTAACCAAGTCTC 1363
QY 375 TCATCTTTAAGACATCATATATATCTGTCATGAAAAAAACATTAAAGGTAAACAATG 434
DB 1364 ACACTTACTACATTAAGATTAATTCATCTGAGAGAAACCTTAACAAATGTAAAGAAATG 1423
QY 435 TATTAAGTTTCACTGACCTTCAAAATTAAGCACACATTAATTAACATCATGTTG 493
DB 1424 TGGCAAGCTTTTAAGCGGTCTCAACCTTACTGAACATAGATTAATTCATATCGAG 1482

RESULT 10
AAS64586
ID AAS64586 standard; CDNA; 3639 BP.

XX
AC AAS64586;
XX
DT 13-FEB-2002 (first entry)
XX
DE DNA encoding novel human diagnostic protein #390.
XX
KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX
OS Homo sapiens.
XX
PN WO200175067-A2.
XX
PD 11-OCT-2001.
XX
PF 30-MAR-2001; 2001WO-US008631.
XX
PR 31-MAR-2000; 2000US-00540217.
PR 23-AUG-2000; 2000US-00649167.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Drmanac RT, Liu C, Tang YT;
XX
DR WPI; 2001-639362/73.
DR P-PSDB; ABO00399.
XX
PT New isolated polynucleotide and encoded polypeptides, useful in
diagnostics, forensics, gene mapping, identification of mutations
responsible for genetic disorders or other traits and to assess
biodiversity.
PT
PS Claim 1; SEQ ID NO 390; 103pp; English.
XX
XX The invention relates to isolated polynucleotide (I) and polypeptide (II)
sequences. (I) is useful as hybridisation probes, polymerase chain
reaction (PCR) primers, oligomers, and for chromosome and gene mapping,
and in recombinant production of (II). The polynucleotides are also used
in diagnostics as expressed sequence tags for identifying expressed
genes. (I) is useful in gene therapy techniques to restore normal

CC activity of (II) or to treat disease states involving (II). (II) is
 CC useful for generating antibodies against it, detecting or quantitating a
 CC polypeptide in tissue, as molecular weight markers and as a food
 CC supplement. (II) and its binding partners are useful in medical imaging
 CC of sites expressing (II). (I) and (II) are useful for treating disorders
 CC involving aberrant protein expression or biological activity. The
 CC polypeptide and polynucleotide sequences have applications in
 CC diagnostics, forensics, gene mapping, identification of mutations
 CC responsible for genetic disorders or other traits to assess biodiversity
 CC and to produce other types of data and products dependent on DNA and
 CC amino acid sequences. AAS64197-AAS94564 represent novel human diagnostic
 CC coding sequences of the invention. Note: The sequence data for this
 CC patent did not appear in the printed specification, but was obtained in
 CC electronic format directly from WIPO at
 CC ftp.wipo.int/pub/published_pct_sequences

XX Sequence 3639 BP; 1403 A; 708 C; 622 G; 906 T; 0 U; 0 Other;

Query Match 7.5%; Score 92.4; DB 5; Length 3639;

Best Local Similarity 54.1%; Pred. No. 1.2e-08;

Matches 236; Conservative 0; Mismatches 191; Indels 9; Gaps 2;

QY 69 AAGTATATTTGACATATGAGGGTGTGATTAAGCTTAATGACCATCATTTATTAGA 128
 DB 738 AGAGAAACCCCTACAAATGTGAGAAATGTGGCAAGCTTTAGCCATTTCAACCTTGC 797
 QY 129 GCACATTTTAAAGAACCCAGTATGATGACCGTAAATGTAACAGTGAACATTTGTA 188
 DB 798 TAAACATTAAGAAATTAATCTGAGAGAAACCCCTAAATGT-----GAAAGATGTGG 851
 QY 189 TAAAGCATTTTGAAGAAATCAATTTGGAACACATTTGTATCAATTCGAAAAAAA 248
 DB 852 CAAGGCTTTAGCGCTTCTCAACCCCTGCTAAACATTAAGAAATTCATCTGAGAGAA 911
 QY 249 ACCATTCATTTGTCAGTGTGTGTAAGGGGTTAATTCGACAAACATTTGAAAAACA 308
 DB 912 ACCCTCAAAATGTAAAGAAATGTGGCAAGCTTTAGCAATTCCTCAACCTTGTATCA 971
 QY 309 TGAATACCCCATTAACAAAGCATTTAATGATCAATTTGAAATTTGCAAGAGCATTTTA 368
 DB 972 TAAGATTAACCTAATCTGAAGAGAAACCCCTCAAAATGTAAAGAAATGTGACAAACCTTTAA 1031
 QY 369 TAAACATCAATC---TTTAAAGACATCATATATTATCTGTCATGAAAAAATTAACGTC 425
 DB 1032 GGCATCTCTCAACCTTAATCAATTAATTAATCATCTGAGAGAAACCTTAACAATG 1091
 QY 426 TAAACATGTAAATTAAGTTTCACTGACCTTCAAAATTAGCAACAATTAATTAACA 485
 DB 1092 TGAAGAAATGTGGCAAGCTTTTATGATCTTCAAAATCTTAATCATATAGTTTATCA 1151
 QY 486 TCATGTGTGATCTCT 501
 DB 1152 TACTGAGAGAGAACT 1167

RESULT 11

ACH87568/C
 ID ACH87568 standard; DNA; 994 BP.

XX ACH87568;
 XX
 DT 29-JUL-2004 (first entry)
 XX Human genome derived single exon probe #20763.
 XX Human; probe; ss; gene expression; single exon probe; microarray;
 KM alternative splicing event; genomic alteration.
 OS Homo sapiens.
 XX US2003194704-A1.
 PN 16-OCT-2003.

XX 03-APR-2002; 2002US-00029386.
 PF 03-APR-2002; 2002US-00029386.
 PR 03-APR-2002; 2002US-00029386.
 XX (PENN/) PENN S G.
 PA (RANK/) RANK D R.
 PA (HANZ/) HANZEL D K.
 XX Penn SG, Rank DR, Hanzel DK;
 PI WPI; 2004-119264/12.
 DR
 XX
 PT New human genome-derived single exon nucleic acid probes useful for human
 PT gene expression analysis, for identifying or characterizing alternative
 PT splicing events, for assessing genomic alterations or as tools for
 PT surveying tissues.
 PT
 PS Claim 1; SEQ ID NO 20763; 80bp; English.

CC The invention relates to a nucleic acid probe for measuring human gene
 CC expression, comprising any of the 27,400 fully defined nucleotide
 CC sequences in the specification, or their complements or fragments, and
 CC encoding at least 8 amino acids of any of the 688 amino acid sequences
 CC fully defined in the specification. The probe is a single exon probe that
 CC hybridizes under high stringency conditions to a nucleic acid molecule
 CC expressed in human cells or tissues. Also included are a spatially-
 CC addressable set of single exon nucleic acid probes for measuring human
 CC gene expression (comprising a plurality of single exon nucleic acid
 CC probes cited above, where each of the plurality of probes is separately
 CC and addressably isolatable or amplifiable from the plurality), a single
 CC exon microarray for measuring human gene expression, a method of
 CC measuring human gene expression, a vector comprising the single exon
 CC probe cited above, an ORF-encoded peptide comprising at least 8
 CC contiguous amino acids of any of the above-mentioned amino acid
 CC sequences (optionally with conservative amino acid substitutions), an
 CC isolated antibody that binds specifically to a peptide cited above,
 CC methods of selling and/or licensing single exon probes or microarrays to
 CC a customer desiring to measure gene expression, a method of providing
 CC human gene expression data by subsequence, and a computer-readable
 CC storage medium which contains a database having a plurality of records
 CC (each record including data on the expression of a single exon probe
 CC cited above. The probe, methods and apparatus are useful in gene
 CC expression analysis. The probes may be used as tools for surveying
 CC tissues to detect the presence of expressed messages that contain their
 CC specific exon, or in constructing genome-derived single exon microarrays.
 CC In addition, the probes are used in identifying and characterizing
 CC alternative splicing events, in detecting and characterizing gross
 CC alterations in the genomic locus that includes their exon, in assessing
 CC smaller genomic alterations, in priming the synthesis of nucleic acids,
 CC or in expressing the ORF-encoded peptide. The present sequence is a human
 CC single exon probe of the invention. Note: The sequence data for this
 CC patent did not form part of the printed specification, but was obtained
 CC in electronic format directly from USPTO at
 CC seqdata.uspto.gov/sequence.html?docid=20030194704

XX Sequence 994 BP; 252 A; 159 C; 203 G; 380 T; 0 U; 0 Other;

Query Match 7.4%; Score 92; DB 12; Length 994;

Best Local Similarity 52.0%; Pred. No. 1.2e-08;

Matches 259; Conservative 0; Mismatches 230; Indels 9; Gaps 2;

QY 7 GAAAGTGAGGAAACCAATGATATGATCTTTAATATCTTCTTCTTCATCAGTCCC 66
 DB 763 GAATGTGGCAAGCCTTCTACATCTTCAACCTTAACATTAAGTAATTCATCTACT 704
 QY 67 AAAAGTATATTTGACATATGAGGGTGTGATTAAGCTTAATGACCATCATTTATTAGA 126
 DB 703 GAGAGAAACCCCTTAATGTGAAGATGTGTAAGCTTTTAAACACCTTCAAGCCCTT 644
 QY 127 GAGCAACATTTAAGAACCCAGATATGATGACCGTATTAATGTACAGTGAAGATTTG 186
 DB 643 ACTACACATTAAGTTCATCTATGTTAAAGAAAAACCTTAACAAATGT-----GAAAGATGT 590

QY	187	CATTAAAGCATTTTTCAGAAAATCAGTTTGGAAACATATTGTATCATTCCGAAAA	246
Db	569	GACAAAGCTTTTAAACCGATTCTCATCTTACTTAAACATTAAGTAATTTCACTTTGGAG	530
QY	247	AAACCATTCATTGTTTCAGTGTGTGTAAAGGGGTTAATTCCTGACACACTTGAANAAG	306
Db	529	AAATCTTACAAATGTGAACAATGTGGCAAAAGGTTTAACTGGTCTTCAACCTTACAAA	470
QY	307	CATGAATACCCATTCACAAAGTCATTTAAATGTACATTTGAAAATTGTCAAGAAGCATTT	366
Db	469	CATTAAGAAATTCATCACTGAGAGAAACCCCTCAAAATGTGAAGAAATGTGGCAAAAGCTTT	410
QY	367	TATTAACAT---CAATCTTTAAGACATCATATATATCTGTTGATGAAAAAAACATTAAAG	423
Db	409	AATGTGTCTTCACACCTTCTACACATTAAGATGATTCATCACTGGAAGAAACCTTCAAA	350
QY	424	TGTAAACATGTATATTAAGTTTTCCTGACCTTCACAAATTAGCACACATTAATTTAAA	483
Db	349	TGTGAAGAAATGTGGCAAAAGCTTTTAAACACTCTCTCAAACTTACTATACATTAAGATAT	290
QY	484	CATCATGTGGGATCTTCTT 501	
Db	289	CATACTGGAGAGAAACCT 272	
RESULT 12			
ID	ADRI4368	standard; DNA; 2873 BP.	
XX	ADRI4368;		
DT	21-OCT-2004	(first entry)	
XX	Human NF-kappaB pathway-associated gene SegID369.		
XX	NF-kappaB pathway; antiinflammatory; cyostatic; hepatotropic; virucide;		
KM	antiarthritic; antirheumatic; gastrointestinal; Gen; antiaesthetic;		
KM	antiartherosclerotic; immunomodulator; cerebroprotective; vasotropic;		
KM	immunosuppressive; vulnary; gene therapy; immune disorder;		
KM	inflammatory disorder; NF-kappaB regulation; cancer; aberrant apoptosis;		
KM	hepatic disorder; Hodgkin's lymphoma; haematopoietic tumour;		
KM	hyper-IgM syndrome; hypohidrotic ectodermal dysplasia;		
KM	X-linked anhidrotic ectodermal dysplasia; immunodeficiency;		
KM	viral infection; HIV-1; HTLV-1; hepatitis B; hepatitis C; EBV; influenza;		
KM	viral replication; host cell survival; evasion of immune response;		
KM	rheumatoid arthritis; inflammatory bowel disease; colitis; asthma;		
KM	atherosclerosis; cachexia; euthyroid sick syndrome; stroke; EAE;		
KM	autoimmune disorder; hyper immune activity;		
KM	aberrant acute phase response; hypercongenital condition; birth defect;		
KM	necrotic lesion; wound; organ transplant rejection;		
KM	aberrant signal transduction; proliferating disorder; cancer;		
KM	HIV propagation; gene; de; human.		
XX			
OS	Homo sapiens.		
XX			
PN	WO2004065577-A2.		
XX			
PD	05-AUG-2004.		
XX			
PF	13-JAN-2004; 2004WO-US000798.		
XX			
PR	14-JAN-2003; 2003US-0440068P.		
XX			
PR	12-MAY-2003; 2003US-0469757P.		
XX			
PA	(BRIM) BRISTOL-MYERS SQUIBB CO.		
XX			
PI	Nadler SG, Neubauer MG, Feder JN, Carman J;		
XX			
DR	WPI; 2004-562168/54.		
XX			
DR	P-PSDB; ADRI4369.		
XX			
PT	New isolated polynucleotides and polypeptides associated with NF-kappaB		

PT pathway, useful for diagnosing, treating, or preventing disorders or
PT diseases associated with NF-kappab pathway.
XX
PS Claim 1; SEQ ID NO 369, 237bp; English.

This invention relates to the novel association of protein sequences (and the genes which encode them) to the NF-kappaB pathway. The invention may be useful for the production of compounds with an anti-inflammatory, cytostatic, hepatotropic, virocidic, antiarthritic, antineumatic, gastroenterostinal-Gen, antistatic, antiatherosclerotic, immunomodulatory, cerebroprotective, vasotropic, immunosuppressive or vulnary activity or for gene therapy. The proteins and nucleotides are useful for diagnosing, preventing, treating, or ameliorating conditions or diseases associated with the NF-kappaB pathway. The condition is an immune disorder, an inflammatory disorder, an inflammatory disorder related to aberrant NF-kappaB regulation, cancer, aberrant apoptosis, hepatic disorders, Hodgkin's lymphomas, haematopoietic tumours, hyper-IgM syndromes, hypohidrotic ectodermal dysplasia, X-linked hidrotic syndromes, ectodermal dysplasia, immunodeficiency, viral infections, HIV-1, hepatitis B, hepatitis C, EBV, influenza, viral replication, host cell survival, evasion of immune responses, rheumatoid arthritis, inflammatory bowel disease, colitis, asthma, atherosclerosis, cachexia, euthyroid sick syndrome, stroke, BAE, autoimmune disorders, disorders related to hyper immune activity, disorders related to aberrant acute phase responses, hypercongenital conditions, birth defects, necrotic lesions, wounds, organ transplant rejection, conditions related to organ transplant rejection, disorders related to aberrant signal transduction, proliferating disorders, cancers and HIV propagation in cells infected with other viruses. The present sequence is that of a human gene which is subject to the novel association with the NF-kappaB pathway of the invention. Note: This sequence does not appear in the specification but was obtained by the indexer from Genbank.

Sequence 2873 BP; 948 A; 564 C; 545 G; 816 T; 0 U; 0 Other;

Query Match 7.48; Score 92; DB 13; Length 2873;

Matches 256; Conservative 0; Mismatches 225; Indels 9; Gaps 2;

Qy 7 GAAAGTGACGAACCAATCGATATCATCTTTAATATCTTCTTCTTCATCAGTCCC 66

Db 1621 GATGTCGCAAAACATTTTGCAATGCTTTCACACCTAACTCAACATAAAGAATCCAACT 1680

67 AAAAGTATATTGACATATGAAGGTGTGATAAAGCCTATAATCGACCATCATTTA 126

Db 1681 AGAGTGAATTTCTACAAATGTGAAGCATATGGAAGAGCCTTTAACTGGTCCCTCAACCTT 1740

QY 127 GAGCAACATTTAAGAACCCACAGTATGATCGACCGTATAATGTACAGTGAGGATTGT 186

Db 1741 AATAACATAAGAGATTCTACTGGAGAAACCCTTACAATGTAA-----GAATGT 179,

QY 187 GATAAGCATTTTTCAGAAATCACATTTGGAACACATATTGTATCACATTCCGAAAA 246

Db 1795 GGCAAGCCTTTAACCAGACCTTATTAGACATAGAGAA TTCATACTGAAGAG 1854

247 AAACCATTCATTGTTCA GTGTGTTAAAGGGTTAATTC TCACACA CACTTGAAGA 306

Db 1855 AAACCTACAATGTGAGATGTGGCAAGCCTTTAACCAAGTCATCGACCCCTTACTACA 1914

307 CATGAAATCACCCTACAAAGTCATTTAAATGTACATTTGAAATTGTCAGAAGCATTT 366

Db 1915 CATAATATAATTCATACTGGGAATTCCCTACAATGTGAGAAATGTTAGAGCTTTT 1974

367 TATAACATCATCTTTAAGACATCATATATTATCTGTT---CATGA AAAACATTACG 423

Db 1975 AACCAAGCCTCAAGCTTACTGAACATAAGTTAATTCATACCGGAGAGAAACGTTATGAA 2034

424 TGTAAACAATGTAATAAGTTTCACTCGACCTTCAAATTAGCACACATAAATTAAA 483

Db 2035 TGTGAAGATGCGCAAGCTTTTAAACCGATCCTCAAACTTACTGAACATAAGTACATT 2094

484 CATCATGGTG 493

—

Db 2095 CATACGTGAG 2104

RESULT 13

ADN99064

ID ADN99064 standard; cDNA; 1533 BP.

XX

AC ADN99064;

XX

DT 29-JUL-2004 (first entry)

XX

DE Novel human cDNA sequence #664.

XX

de; gene; anti-inflammatory; dermatological; neuroprotective; immunomodulator; antibacterial; virucide; antiproliferative; cytostatic; gene therapy; vaccine; inflammatory; CNS; immune disorder; cancer; psoriasis; diabetes; early aging; hormonal imbalance; ischemic heart disease; ulcerative colitis.

XX

OS Homo sapiens.

XX

PN WO2004038003-A2.

XX

PD 06-MAY-2004.

XX

PF 24-OCT-2003; 2003WO-US033947.

XX

PR 25-OCT-2002; 2002US-0421061P.

PR 25-OCT-2002; 2002US-0421080P.

PR 25-OCT-2002; 2002US-0421552P.

PR 25-OCT-2002; 2002US-0421614P.

PR 30-OCT-2002; 2002US-0422177P.

PR 30-OCT-2002; 2002US-0422178P.

PR 15-NOV-2002; 2002US-0426355P.

PR 15-NOV-2002; 2002US-0426384P.

PR 15-NOV-2002; 2002US-0426394P.

PR 15-NOV-2002; 2002US-0426430P.

PR 15-NOV-2002; 2002US-0426916P.

PR 27-NOV-2002; 2002US-0429224P.

PR 27-NOV-2002; 2002US-0429275P.

PR 27-NOV-2002; 2002US-0429302P.

PR 27-NOV-2002; 2002US-0429326P.

PR 27-NOV-2002; 2002US-0429651P.

PR 04-DEC-2002; 2002US-0430645P.

PR 04-DEC-2002; 2002US-0430651P.

PR 04-DEC-2002; 2002US-0430657P.

PR 04-DEC-2002; 2002US-0430663P.

PR 04-DEC-2002; 2002US-0430668P.

PR 04-DEC-2002; 2002US-0430684P.

PR 05-DEC-2002; 2002US-0430937P.

PR 05-DEC-2002; 2002US-0430965P.

PR 12-DEC-2002; 2002US-0431458P.

PR 12-DEC-2002; 2002US-0433251P.

PR 13-DEC-2002; 2002US-0433316P.

PR 13-DEC-2002; 2002US-0433318P.

PR 23-DEC-2002; 2002US-0436236P.

PR 03-JAN-2003; 2003US-0437914P.

PR 17-JAN-2003; 2003US-0440820P.

PR 17-JAN-2003; 2003US-0440821P.

PR 18-APR-2003; 2003US-0463760P.

PR 18-APR-2003; 2003US-0463768P.

PR 18-APR-2003; 2003US-0463768P.

PR 18-APR-2003; 2003US-0463768P.

PR 18-APR-2003; 2003US-0463768P.

PR 02-MAY-2003; 2003US-0467199P.

PR 02-MAY-2003; 2003US-0467201P.

PR 02-MAY-2003; 2003US-0467203P.

PR 02-MAY-2003; 2003US-0467230P.

PR 19-MAY-2003; 2003US-0471306P.

PR 19-MAY-2003; 2003US-0471336P.

PR 22-MAY-2003; 2003US-0472420P.

PR 22-MAY-2003; 2003US-0472430P.

PR 09-JUN-2003; 2003US-0476609P.

PR 09-JUN-2003; 2003US-0476621P.

PR 09-JUN-2003; 2003US-0476632P.

PR 09-JUN-2003; 2003US-0476641P.

PR 08-JUL-2003; 2003US-0485217P.

PR 08-JUL-2003; 2003US-0485218P.

PR 08-JUL-2003; 2003US-0485223P.

PR 08-JUL-2003; 2003US-0485224P.

PR 08-JUL-2003; 2003US-0485259P.

PR 14-JUL-2003; 2003US-0486446P.

PR 14-JUL-2003; 2003US-0486480P.

PR 15-JUL-2003; 2003US-0486891P.

PR 15-JUL-2003; 2003US-0486960P.

PR 08-AUG-2003; 2003US-0493341P.

PR 08-AUG-2003; 2003US-0493370P.

PR 08-AUG-2003; 2003US-0493573P.

PR 08-AUG-2003; 2003US-0493577P.

XX

PA (FIVE-) FIVE PRIME THERAPEUTICS INC.

XX

PI Williams LT, Chu K, Lee E, Heatir K, Beaurang PA, Behrens D,

PI Halendeck RF, Kochakota S, Lin H, Linemann T, Pierce K, Wang Y,

PI Wong JGP, Wu G, Zhang H, Zeng C;

DR WPI; 2004-36511/34.

DR P-PsDB; ADN9848.

XX

PS Claim 1; SEQ ID NO 664; 532bp; English.

XX

CC The invention relates to a nucleic acid molecule comprising a

CC polynucleotide sequence or its complement that encodes a polypeptide. The

CC nucleic acid is useful in preparing a composition for treating or

CC preventing inflammatory, CNS, immune, bacterial or viral disorder.

CC cancer, psoriasis, diabetes, early aging, hormonal imbalance, ischemic

CC heart disease or ulcerative colitis. This sequence corresponds to a

CC nucleic acid of the invention.

XX

SQ Sequence 1533 BP; 614 A; 280 C; 251 G; 388 T; 0 U; 0 Other;

Query Match 7.4%; Score 91.8; DB 12; Length 1533;

Best Local Similarity 52.1%; Pred. No. 1.4e-08;

Matches 257; Conservative 0; Mismatches 227; Indels 9; Gaps 2;

QY 11 GTGACGAACCAATGATATATCTTTAATATCTTCTTCTTATCATCGCCCAAA 70

DB 305 GTGGCAATCATTTTGCATGCTTTCACAATTAATCAACATTAAGAAATTCATCTAGAG 364

QY 71 AGTATATTTGCACATATGAAGGGTGTATTAAGCCTATATCGACCATCTTTAGAGC 130

DB 365 AGTATCTTACAAATGTGAAGATGTGTAAACCTTTAACTGTCTCAACCTTAACCTTA 424

QY 131 AACATTGAAGAACCCACAGTATGATGACCGTATTAATGTATACAGTGAAGCATTTGATA 190

DB 425 AACATAAGATTAATCTACTGAGGAAAAAACCCTACAAATGT-----GAAGATGTGGCA 478

QY 191 AAGCATTTTTCAGAAATATCATTTGGAAACATATTAATGTATGATCATTCGAGAAAAAAG 250

DB 479 AAGCTTTTAAACGGTCTCAAAATCTTAATTAATTAATTAATTAATTAATTAATTAATTA 538

QY 251 CATTCATTTTTCAGTGTGTGTGAAGGGGTATTTCTGACAAACCTGAAAAAGCATG 310

DB 539 CTTACAAATGTGAAGATGTGGCAAAAGCTTTTAAACGGTCTCAACCTTAAACATA 598

QY 311 AATACCCATACAAATGATATTAATGTATGATTTGAAATTTGTCAAGAGCATTTTATA 370

DB 599 AAAGAATTCATACGAAGAAACCTTACAAATGTATGAGATGTGCAAGGCTTTAACCC 658

QY 371 AACATCATCTT---AAGCATCATATATTAATCTGTATGAAAAAAGATTAACGTGTA 427

```

Db      659 AATTCTCGATTCCTTATATAAACAATTAAGAGATTCAATGGAAGATTAACCTTCAATGTG 718
Oy      428 AACAAATGATATTAAGTTTCACTCGACCTTCAAAATTTAGCAGACATAATTAATAATC 487
Db      719 AAGATGTGTGGCAAGGCTTTAGAGTATTTCTCTTAAAAAACAATAAGATATATCCATA 778
Oy      488 ATGTGTGATCTCC 500
Db      779 CTGGGGAATAAC 791

RESULT 14
AD000633
ID      AD000633 standard; cDNA; 1533 BP.
XX
XX
AC      AD000633;
XX
DT      29-JUL-2004 (first entry)
XX
DE      Novel human cDNA sequence #1448.
XX
KW      de; gene; anti-inflammatory; dermatological; neuroprotective;
KW      immunomodulator; antibacterial; virucide; antiparasitic; cytostatic;
KW      gene therapy; vaccine; inflammatory; CNS; immune disorder; cancer;
KW      psoriasis; diabetes; early aging; hormonal imbalance;
KW      ischemic heart disease; ulcerative colitis.
XX
OS      Homo sapiens.
XX
PN      WO2004038003-A2.
XX
PD      06-MAY-2004.
XX
PF      24-OCT-2003; 2003WO-US033947.
XX
PR      25-OCT-2002; 2002US-0421061P.
PR      25-OCT-2002; 2002US-0421080P.
PR      25-OCT-2002; 2002US-0421552P.
PR      25-OCT-2002; 2002US-0421614P.
PR      30-OCT-2002; 2002US-0422177P.
PR      30-OCT-2002; 2002US-0422178P.
PR      15-NOV-2002; 2002US-0426355P.
PR      15-NOV-2002; 2002US-0426384P.
PR      15-NOV-2002; 2002US-0426394P.
PR      15-NOV-2002; 2002US-0426430P.
PR      15-NOV-2002; 2002US-0426916P.
PR      27-NOV-2002; 2002US-0429244P.
PR      27-NOV-2002; 2002US-0429275P.
PR      27-NOV-2002; 2002US-0429302P.
PR      27-NOV-2002; 2002US-0429326P.
PR      27-NOV-2002; 2002US-0429651P.
PR      04-DEC-2002; 2002US-0430645P.
PR      04-DEC-2002; 2002US-0430651P.
PR      04-DEC-2002; 2002US-0430657P.
PR      04-DEC-2002; 2002US-0430663P.
PR      04-DEC-2002; 2002US-0430668P.
PR      04-DEC-2002; 2002US-0430684P.
PR      05-DEC-2002; 2002US-0430937P.
PR      05-DEC-2002; 2002US-0430965P.
PR      12-DEC-2002; 2002US-0431458P.
PR      12-DEC-2002; 2002US-0433251P.
PR      12-DEC-2002; 2002US-0433500P.
PR      13-DEC-2002; 2002US-0433316P.
PR      13-DEC-2002; 2002US-0433318P.
PR      23-DEC-2002; 2002US-0436236P.
PR      03-JAN-2003; 2003US-0437914P.
PR      17-JAN-2003; 2003US-0440820P.
PR      17-JAN-2003; 2003US-0440821P.
PR      18-APR-2003; 2003US-0463700P.
PR      18-APR-2003; 2003US-0463708P.
PR      18-APR-2003; 2003US-0463716P.
PR      18-APR-2003; 2003US-0463732P.

```

```

PR      02-MAY-2003; 2003US-0467199P.
PR      02-MAY-2003; 2003US-0467201P.
PR      02-MAY-2003; 2003US-0467203P.
PR      02-MAY-2003; 2003US-0467203P.
PR      19-MAY-2003; 2003US-0471306P.
PR      19-MAY-2003; 2003US-0471336P.
PR      22-MAY-2003; 2003US-0472420P.
PR      22-MAY-2003; 2003US-0472430P.
PR      09-JUN-2003; 2003US-0476609P.
PR      09-JUN-2003; 2003US-0476621P.
PR      09-JUN-2003; 2003US-0476632P.
PR      09-JUN-2003; 2003US-0476641P.
PR      08-JUL-2003; 2003US-0485217P.
PR      08-JUL-2003; 2003US-0485218P.
PR      08-JUL-2003; 2003US-0485223P.
PR      08-JUL-2003; 2003US-0485224P.
PR      08-JUL-2003; 2003US-0485325P.
PR      08-JUL-2003; 2003US-0485359P.
PR      14-JUL-2003; 2003US-0486446P.
PR      14-JUL-2003; 2003US-0486480P.
PR      15-JUL-2003; 2003US-0486891P.
PR      15-JUL-2003; 2003US-0486960P.
PR      08-AUG-2003; 2003US-0493341P.
PR      08-AUG-2003; 2003US-0493370P.
PR      08-AUG-2003; 2003US-0493573P.
PR      08-AUG-2003; 2003US-0493577P.

(FIVE-) FIVE PRIME THERAPEUTICS INC.
XX
XX      William LT, Chu K, Lee E, Heistr K, Beaurang PA, Behrens D;
PI      Halenbeck RP, Kochakota S, Lin H, Linnemann T, Pierce K, Wang Y;
PI      Wong JG, Wu G, Zhang H, Zeng C;
XX
DR      MPI; 2004-365511/34.
XX
DR      P-PSDB; ADN9848.
XX
PT      New nucleic acid molecules, useful in preparing a composition for
PT      treating or preventing e.g. inflammatory, CNS, bacterial or viral
PT      disorders, cancer, psoriasis, diabetes, ischemic heart disease or
XX      ulcerative colitis.
XX
PS      Claim 1, SEQ ID NO 2232, 532pp; English.
XX
XX      The invention relates to a nucleic acid molecule comprising a
XX      polynucleotide sequence or its complement that encodes a polypeptide. The
XX      nucleic acid is useful in preparing a composition for treating or
XX      preventing inflammatory, CNS, immune, bacterial or viral disorder,
XX      cancer, psoriasis, diabetes, early aging, hormonal imbalance, ischemic
XX      heart disease or ulcerative colitis. This sequence corresponds to a
XX      nucleic acid of the invention.
XX
SQ      Sequence 1533 BP; 614 A; 280 C; 251 G; 388 T; 0 U; 0 Other;

Query Match      7.4%; Score 91.8; DB 12; Length 1533;
Best Local Similarity 52.1%; Pred. No. 1.4e-08;
Matches 257; Conservative 0; Mismatches 227; Indels 9; Gaps 2;

Oy      11 GTGAGAAACCAATCGATATCATCTTTAATATCTCTTCTTCATCAGGCCCAAA 70
Db      305 GTGGCAATCATTTTGCATGCTTTCAACATTAACACATAGAAATTCATCTAGAG 364
Oy      71 AGTATATTTGCACATATGAGGGGTGATTAAGCCTATTAATGACCATATTATAGAC 130
Db      365 AGTATTTCTTACAAATGTGAAGATGTGTAAGCCTTTAATCTGTCCTAACCCCTACTA 424
Oy      131 AACATTAAAGAACCAAGTAATGATCGACCGTATTAATGTACAGTGACGATTTGTGATA 190
Db      425 AACATTAAGATTAATCTATCTGAGAAAAACCTTACAAATGT-----GAAGATGTGGCA 478
Oy      191 AAGCATTTTTCAGAAATTCATCTTTGAAAGACATATTTGATACATTCGAAAAAANAAC 250
Db      479 AAGCTTTAACCGGTCTCAAAATCTTACTTAACATTAATAATTAATCTACTGAGAGAAAC 538

```

QY 251 CATTCATTGTTGATGTTGTTAAAGGGTTAATTCGACAACTTGAAAAACATG 310
 DB 539 CCTACAAATGTGAAGATGTGGCAAGCTTTAAACGGTCTCAACCTTACTAATACATA 598
 QY 311 AATTCACCCATACAAAGTCAATTTAAATGTCATTGAAAAATGTCAGAAAGCATTTATA 370
 DB 599 AAAGAAATCTATACAGAAAGAAACCCCTACAAAGTGAAGATGTGGCAAGCTTTAAAC 658
 QY 371 AACATCAATCTT---AAGACATCATATATATCTGTTGATGAAAAACATTAACGTGA 427
 DB 659 AGTTCGATTTCTTAATTAACATTAAGAAATTCATATGAAAGTAAACCTTACAAATTTG 718
 QY 428 AACATGTAATTAAGTTTCTGACCTGACCTTCAAAATTAAGCAACATTAATTAATTAATC 487
 DB 719 AAGAAATGTGCAAGCTTTAGAGTATCTCAATTTCTTAATAAAACATTAATATCATATA 778
 QY 488 ATGTGATCTCC 500
 DB 779 CTGGGAAAAAAC 791

RESULT 15

ADB63624 standard, cDNA; 2110 BP.
 ID ADB63624 standard, cDNA; 2110 BP.
 XX
 AC ADB63624;
 XX
 DT 04-DEC-2003 (first entry)
 XX
 DE Human cDNA encoding clone THYM20071120.
 XX
 KM Human; ss; gene; pharmaceutical; diagnostic; gene therapy;
 KM tissue regeneration; cell regeneration; membrane protein;
 KM signal transduction-related protein; transcription-related protein;
 KM osteoporosis; neurological disease; cancer; tumour.
 OS Homo sapiens.
 XX
 FH Key Location/Qualifiers
 FT CDS 170..1792
 FT /tag= a
 FT /product= "Clone THYM20071120 protein"

EP1308459-A2.
 PD 07-MAY-2003.
 XX
 PF 28-MAR-2002; 2002EP-00007401.
 XX
 PR 05-NOV-2001; 2001JP-00379298.
 PR 25-JAN-2002; 2002US-00350978.
 XX
 PA (HELI-) HELIX RES INST.
 PA (REAS-) RES ASSOC BIOTECHNOLOGY.
 XX
 PI Isegaki T, Sugiyama T, Otsuki T, Wakamatsu A, Sato H, Ishii S,
 PI Yamamoto J, Isono Y, Hio Y, Otsuka K, Negai K, Irie R, Tamechika I,
 PI Seki N, Yoshihawa T, Otsuka M, Negahari K, Masuno Y;
 XX
 DR MPI; 2003-450961/43.
 DR P-PSDB; ADB65594.
 XX
 PT New polynucleotides and polypeptides, useful for developing a diagnostic
 PT marker or medicines for regulation of their expression and activity, or
 PT as targets of gene therapy.
 XX
 PS Claim 1; Page; 222pp; English.
 XX
 CC The invention discloses a polynucleotide comprising a sequence selected
 CC from 1970 fully defined nucleotide sequences which encode novel
 CC polypeptides. Also claimed is a polypeptide encoded by the polynucleotide
 CC or its partial peptide, an antibody binding to the polypeptide or peptide
 CC of the polynucleotide, immunologically assaying the polypeptide or

CC peptide of the polynucleotide by contacting the polypeptide or peptide
 CC with the antibody of the encoded protein, and observing the binding
 CC between the two, a transformant carrying the polynucleotide in an
 CC expressible manner and an antisense polynucleotide. The oligonucleotide
 CC is useful as a primer for synthesizing the polynucleotide, and as a probe
 CC for detecting the polynucleotide. The polynucleotides and encoded
 CC proteins are useful as pharmaceutical agents and many disease-related
 CC genes may be included in them, for developing a diagnostic marker or
 CC medicines for regulation of their expression and activity, or as targets
 CC of gene therapy. The genes are involved in tissue and/or cell
 CC regeneration. Membrane proteins, signal transduction-related proteins,
 CC transcription-related proteins, disease-related proteins and genes
 CC encoding them can be used as indicators for diseases (e.g. osteoporosis,
 CC neurological diseases, cancer, tumours. The cDNA may be used to regulate
 CC the activity or expression of the encoded protein to treat diseases. The
 CC sequence presented is a cDNA of the invention. Note: Some of the sequence
 CC data for this patent is not represented in the printed specification, but
 CC is based on sequence information supplied by the European Patent Office.
 XX
 SQ Sequence 2110 BP; 784 A; 394 C; 386 G; 546 T; 0 U; 0 Other;

Query Match 7.4%; Score 91.8; DB 10; Length 2110;
 Best Local Similarity 51.7%; Pred. No. 1.5e-08;
 Matches 314; Conservative 0; Mismatches 277; Indels 16; Gaps 4;

QY 61 CGTCCCAAAAGATATTTGACATATGAAGGGTGTGATTAAGCTTAAATGACCATCA 120
 DB 1253 CATACCGAAGAGAAACCTTACAAATGTAAAGATGTGGCAAGCTTTAAACCTCTCA 1312
 QY 121 TTATTAAGCAACATTTAAGAACCCACAGTAATGATGACCGTATTAATGTACAGTGAC 180
 DB 1313 GCCCTTACATACATTAAGAAATTCACCTGAGAGAAACCCATCAATGT-----GAA 1366
 QY 181 GATTGTGATTAAGCAATTTTCAAGAAATCAATTTGAAACATATTTGTATACATTC 240
 DB 1367 GAATGTGGCAAGCTTTAAGCCATCTCAAACTTAAGTGAACATTAAGAAACCTTCACTACT 1426
 QY 241 GAAAGAAAACCATTCATGTTTCAGTGTGTGTAAGGGGTATATCTGCAACAACCTTG 300
 DB 1427 GGAAGAAAACCTTACAAATGTGAAGATGTGGCAAGCTTTTATCCAAATCTTCAAACTT 1486
 QY 301 AAAAGCATGAATCAACCATCAAAAGTCAATTTAATGTACATTTGAAATTTGTCAAGAA 360
 DB 1487 ACTGAACATTAAGAAATTTCTATCTGAGAGATACCTTACAGTGTGAAGATGTGGCAA 1546
 QY 361 GCATTTTATTAACATCAATCTTTA---AAGCATCATATATATCTGTCATGAAGAAACA 417
 DB 1547 GCTTTTAAACACTCTCATCTCTTACTACATTAAGAAATTCATCTGGGAGAAACCC 1606
 QY 418 TTAACTGTAAACAATGTAAATTAAGTTTCACCTGACCTTCAAAATTAAGCAACAATTA 477
 DB 1607 TACAAATGTGAAGATGTGGCAAGCTTTTAAAGCCGATCTCAAAACTTACTGAACATTA 1666
 QY 478 TTAAACATCATGATG-GATCTCTGCTTATCATATGATGATCTCTGTTGTTTAAAAA 536
 DB 1667 ATATATTCATACCTGAGAGAAACCTTAAATGTGAAGATGTGACAAAGCTTTTAAACAA 1726
 QY 537 TTTCCAAATGTTGATGATTTACATTTCAATTAATAACAATGATCAAAACTTAAATG 596
 DB 1727 TCTGCAAAACCT-----TACTAAACATTAAGAAATCATATCTGGAAGAACTACAGAAC 1780
 QY 597 TCTTAATGTGTAAGATGTTGTTGGAAAAAGTTTATCTTCAATATGTTAAGTCA 656
 DB 1781 TGAATGTGTGATTAATGATTTTGAACAACCTTCAAAATTTTCTAAATTAAGAAATCA 1840
 QY 657 TGATGAT 663
 DB 1841 TACTGCT 1847

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd

OM nucleic - nucleic search, using bw model

Run on: February 9, 2005, 01:51:00 ; Search time 223.728 Seconds
(without alignments)
9039.698 Million cell updates/sec

Title:	US-09-831-804-1_COPY_720_1955
Perfect score:	1236
Sequence:	1 atgagtgaagtgcgaac.....aacatcagtgtattctcga 1236

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1202784 seqs, 818138359 residues

Total number of hits satisfying chosen parameters: 2405568

```
Minimum DB seq length: 0
Maximum DB seq length: 20000000000
```

Maximum Match 100% Listing First 45 summaries

```

Database :
Issued Patents NA:*
1: /cgn2_6/prodata/1/ina/5A_COMB_seg.*
2: /cgn2_6/prodata/1/ina/5B_COMB_seg.*
3: /cgn2_6/prodata/1/ina/6A_COMB_seg.*
4: /cgn2_6/prodata/1/ina/6B_COMB_seg.*
5: /cgn2_6/prodata/1/ina/PTUSS_COMB_seg.*
6: /cgn2_6/prodata/1/ina/backfill_seg.*

```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	581.8	47.1	627	4	US-09-248-796A-5184	Sequence 5184, App
2	391	31.6	462	4	US-09-248-796A-5186	Sequence 5186, App
3	96.6	7.8	2320	4	US-09-016-434-1054	Sequence 1054, App
4	90.8	7.3	3798	4	US-09-949-016-4204	Sequence 4204, App
5	90.8	7.3	158942	4	US-09-949-016-12227	Sequence 12227, App
6	90.8	7.3	158950	4	US-09-949-016-15946	Sequence 15946, App
7	89.2	7.2	3839	4	US-09-949-016-4885	Sequence 4885, App
8	86.8	7.0	601	4	US-09-949-016-77359	Sequence 27359, App
9	86.8	7.0	601	4	US-09-949-016-149638	Sequence 149638, App
10	86.2	7.0	601	4	US-09-949-016-44799	Sequence 44799, App
11	86.2	7.0	45138	4	US-09-949-016-3307	Sequence 13027, App
12	85.6	6.9	601	4	US-09-949-016-77360	Sequence 27360, App
13	85.6	6.9	601	4	US-09-949-016-149637	Sequence 149637, App
14	83.2	6.7	601	4	US-09-949-016-17361	Sequence 27361, App
15	83.2	6.7	601	4	US-09-949-016-149636	Sequence 149636, App
16	81.6	6.6	472	4	US-09-513-999C-12522	Sequence 12522, App
17	80.2	6.5	2804	4	US-09-949-016-2278	Sequence 2278, App
18	80.2	6.5	22294	4	US-09-949-016-14200	Sequence 14200, App
19	79.4	6.4	1030	4	US-09-976-594-170	Sequence 370, App
20	79.2	6.4	2982	4	US-09-949-016-1184	Sequence 4184, App
21	79.2	6.4	41435	4	US-09-949-016-15926	Sequence 15926, App
22	78.2	6.3	3400	4	US-09-799-451-136	Sequence 136, App
23	77.2	6.2	601	4	US-09-949-016-149473	Sequence 149473, App
24	76	6.1	601	4	US-09-949-016-149474	Sequence 149474, App
25	76	6.1	2501	4	US-09-949-016-5363	Sequence 5363, App
26	76	6.1	2504	4	US-09-949-016-487	Sequence 487, App
27	76	6.1	2671	4	US-09-949-016-12229	Sequence 12229, App

ALIGNMENTS

28	76	1	29621	4	US-09-949-016-617105	Sequence 17105, A
29	74.4	6.0	2143	4	US-09-881-578A-3	Sequence 3, Appl1
30	73	5.9	2925	4	US-09-620-312D-163	Sequence 193, App
31	71.2	5.8	929	4	US-09-976-584-339	Sequence 969, App
32	69.6	5.6	2402	4	US-09-949-016-1710	Sequence 1710, App
33	68.7	5.6	2407	4	US-09-949-016-189	Sequence 489, App
34	69.6	5.6	30867	4	US-09-949-016-12231	Sequence 12231, App
35	69.6	5.6	30868	4	US-09-949-016-611452	Sequence 13452, A
36	68	5.5	28981	4	US-09-949-016-1115	Sequence 1115, App
37	67.6	5.5	601	4	US-09-949-016-639838	Sequence 39838, A
38	67.6	5.5	601	4	US-09-949-016-160861	Sequence 160861, A
39	66.8	5.4	18773	4	US-09-949-016-614164	Sequence 14164, A
40	66.4	5.4	2891	4	US-09-949-016-4503	Sequence 4503, App
41	66.4	5.4	21360	4	US-09-949-016-12857	Sequence 12857, A
42	66.4	5.4	21361	4	US-09-949-016-16245	Sequence 16245, A
43	66.2	5.4	2752	4	US-09-949-016-6563	Sequence 3563, App
44	65.2	5.4	18824	4	US-09-949-016-15305	Sequence 15305, A
45	65.4	5.3	2274	4	US-09-620-312D-154	Sequence 154, App

```

, RESULT 1
, US-09-248-796A-5184
, Sequence 5184, Application US/09248796A
, Patent No. 6747137
, GENERAL INFORMATION:
, APPLICANT: Keith Weinstein et al
, TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBICA
, FILE OF INVENTION: FOR DIAGNOSTICS AND THERAPEUTICS
, FILE REFERENCE: 107196.132
, CURRENT APPLICATION NUMBER: US/09/248,796A
, CURRENT FILING DATE: 1999-02-12
, PRIOR APPLICATION NUMBER: US 60/074,725
, PRIOR FILING DATE: 1998-02-13
, PRIOR APPLICATION NUMBER: US 60/096,409
, PRIOR FILING DATE: 1998-08-13
, NUMBER OF SEQ ID NOS: 28208
, SEQ ID NO 5184
, LENGTH: 627
, TYPE: DNA
, ORGANISM: Candida albicans
, US-09-248-796A-5184

```

Query Match	Best Local Similarity	97.1%;	Score 581.8;	DB 4;	Length 627;
Matches	605;	Conservative	0;	Mismatches	7;
				Indels	9;
				Gaps	1
QY	616	TGTGTGGGAAAAAGGTTTATCTTCACATATGTTAACTCATGATGATTCACCATGATC	675		
DB	13	TGTTTGGGAAAAAAGGTTTATCTTCACATATGTTAAAGTCATGATGATTTTACCATGATC	72		
QY	676	AAAAATGAGCTTGATGATTATTGCGATGTGGGAAATTTGCCAAGAAAAATGAAATTAGTT	735		
DB	73	AAAAATGAGACTTGATGATTATTGCGATGTGGGAAATTTGCCAAGAAAAATGAAATTAGTT	132		
QY	736	GAACTTATTAATATCTTCATGATGATGATTAATCCCTGATGATTTATTTAAAGAACTGAA	795		
DB	133	GAACTTATTAATATCTTCATGATGATGATTAATCCCGATGATTTATTTAAAGAACTGAA	192		
QY	796	GTGAAAAAATATGAGAACCTTATTGATCAAGAGATGGAATTTAAATTTTGATGAAATTA	855		
DB	193	GTGAAAAAATATGAGAACCTTATTGATCAAGAGATGGAATTTAAATTTTGATGAAATTA	252		
QY	856	GAAAACAGGAATTTAAAAGTGAAGAGATGAAGAGATGAAGAGATGATGCTAGATGA	915		
DB	253	GAAAACAGGAATTTAAAAGTGAAGAGATGAAGAGATGATGCTAGATGA	303		
QY	916	AAAAAGATGATGTTAGTCAAGCTCAATGTCAGTCAAGATCAATTAATCAATTAATCAATTA	975		
DB	304	AAAAAGATGATGTTAGTCAAGCTCAATGTCAGTCAAGATCAATTAATCAATTAATCAATTA	363		

QY 976 GCTTCTTGGAGGTTCAAGAGTGTCTTCTAACTTATCTGATATGTGGAAAGATC 1035
DB 364 GCTTCTTGGAGGTTCAAGAGTGTCTTCTAACTTATCTGATATGTGGAAAGATC 423
QY 1036 AATTGCTTAGAATAATGTGTAGTATGTTTCTAGAGATATATATTAGTGCACAT 1095
DB 424 AATTGCTTAGAATAATGTGTAGTATGTTTCTAGAGATATATATTAGTGCACAT 483
QY 1096 TTGAAATGCGATGATATATTTTACAAAGATTGACTCATTTCTTAATATGATAGAAAA 1155
DB 484 TTGAAATGCGATGATATATTTTACAAAGATTGACTCATTTCTTAATATGATAGAAAA 543
QY 1156 GAAAGAACTCCAGAGGTGAACCATTTGGTTAAAAAACCAGATGATTTATGCCAAAT 1215
DB 544 GAAAGAACTCCAGAGGTGAACCATTTGGTTAAAAAACCAGATGATTTATGCCAAAT 603
QY 1216 GAAACATCAGTGTCTTCTCGA 1236
DB 604 GAAACATCAGTGTCTTCTCGA 624

RESULT 2

US-09-248-796A-5186
Sequence 5186, Application US/09248796A
Patent No. 6747137
GENERAL INFORMATION:
APPLICANT: Keith weinsock et al
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBICAN
FILE REFERENCE: 107196.132
CURRENT APPLICATION NUMBER: US/09/248,796A
CURRENT FILING DATE: 1999-02-12
PRIOR APPLICATION NUMBER: US 60/074,725
PRIOR FILING DATE: 1998-02-13
PRIOR APPLICATION NUMBER: US 60/096,409
PRIOR FILING DATE: 1998-08-13
NUMBER OF SEQ ID NOS: 28208
SEQ ID NO 5186
LENGTH: 462
TYPE: DNA
ORGANISM: Candida albicans
US-09-248-796A-5186

Query Match 31.6%; Score 391; DB 4; Length 462;
Best Local Similarity 98.5%; Pred. No. 4.1e-80;
Matches 405; Conservative 0; Mismatches 5; Indels 1; Gaps 1;

QY 1 ATGAGTGAAGTGAAGCAAAACCAATGATATCATCTTTAATATCTTCTTCTTCATCA 60
DB 52 ATGAGTGAAGTGAAGCAAAACCAATGATATCATCTTTAATATCTTCTTCTTCATCA 111
QY 61 CGTCCCAAAAAGTATATTTGACATATGAAGGGTGTATGAAGCCTATATGACCATCA 120
DB 112 CGTCCCAAAAAGTATATTTGACATATGAAGGGTGTATGAAGCCTATATGACCATCA 171
QY 121 TTATTTAGCAACATTTAAGAACCCAGATATGATGACCGTATTAATGTACATGAC 180
DB 172 TTATTTAGCAACATTTAAGAACCCAGATATGATGACCGTATTAATGTACATGAC 231
QY 181 GATTGTGATTAAGCATTTTTCAGAAAATCATTTTGGAAACACATTTGTATCATCTCC 240
DB 232 GATTGTGATTAAGCATTTTTCAGAAAATCATTTTGGAAACACATTTGTATCATCTCC 291
QY 241 GAAAAAATCAATTCATTTGAGTGTGTGTTAAAGGGGTTAATCTCGAACAACCTTG 300
DB 292 GAAAAAATCAATTCATTTGAGTGTGTGTTAAAGGGGTTAATCTCGAACAACCTTG 351
QY 301 AAAAGCATGAATCAACCATCAAGTCAATTTAAATGTACATTTGAAATGTCAAGAA 360
DB 352 AAAAGCATGAATCAACCATCAAGTCAATTTAAATGTACATTTGAAATGTCAAGAA 411
QY 361 GCATTTTAT-AAAATCAATCTTTAAGCATCATATATATCTGTTCATGA 410

DB 412 ACATTTTATAAACAATCAATCTTTAAGACATCATATATATCTGTTCATGA 462

RESULT 3

US-09-016-434-1054
Sequence 1054, Application US/09016434
Patent No. 6500938
GENERAL INFORMATION:
APPLICANT: Janice Au-Young
APPLICANT: Jeffrey J. Seilhamer
TITLE OF INVENTION: COMPOSITION FOR THE DETECTION OF SIGNALING
TITLE OF INVENTION: PATHWAY GENE EXPRESSION
NUMBER OF SEQUENCES: 1490
CORRESPONDENCE ADDRESS:
ADDRESSEE: INCYTE PHARMACEUTICALS, INC.
STREET: 3174 PORTER DRIVE
CITY: PALO ALTO
STATE: CALIFORNIA
COUNTRY: USA
ZIP: 94304
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Word Perfect 6.1 for Windows/MS-DOS 6.2
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/016,434
FILING DATE: HEREWITH
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Zeller, Karen J.
REGISTRATION NUMBER: 37,071
REFERENCE/DOCKET NUMBER: PA-0002 US
TELEPHONE: (650) 855-0555
TELEFAX: (650) 845-4166
INFORMATION FOR SEQ ID NO: 1054:
SEQUENCE CHARACTERISTICS:
LENGTH: 2320 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
LIBRARY: GENBANK
CLONE: g1017721
US-09-016-434-1054

Query Match 7.8%; Score 96.6; DB 4; Length 2320;
Best Local Similarity 51.3%; Pred. No. 1.4e-12;
Matches 306; Conservative 0; Mismatches 279; Indels 12; Gaps 3;

QY 69 AAGATATATTTGACATATGAAGGGTGTATGAAGCCTATATATGACCATCATTTATGA 128
DB 964 AAGAAACCTTCAATGTAAATGTGTAAAGCTTTTAAACGATCTTCAACCTTAC 1023
QY 129 GCAACATTTAAGAACCCAGATATGATGACCGTATTAATGTACATGAGACATTTGCA 188
DB 1024 TACCATATGAATAATTTCACTGAGAGAAACCTTACAAATGT-----GAAGAATGTG 1077
QY 189 TAAAGCATTTTCAAAAAATCAATTTGGAACACATATTTGTATCATCTTCGAAAAAAA 248
DB 1078 CAAGCCTTTAAGCAGTCTCTCAACCTTATCACTAAGATATATTTCACTGAGAGAA 1137
QY 249 ACCATTCAATTTGTCAGTGTGTGTTAAAGGGGTTAATCTCGAACAACCTTGAAGAACA 308
DB 1138 ACCCTAACAATGTAAATAATGTGAAAGCCCTTTAACAAGTGTGACACCTTACACACA 1197
QY 309 TGAATTCACCATCAAGATCATTTAAATGTACATTTGAAATTTGTCAAGAAACATTTTA 368

Db 1198 TGAAGTAATTCATACGAGAGAAACCTTCAAAATGTGAAAAATGTGAAAAAGCTTTAA 1257
Qy 369 TAAACA---TCAATCTTTAAGACATCATATATATCTGTTTCATGAAAAAATTAACG 425
Db 1258 TCATTTCTCACACCTTCTACATTAAGATTAATTCATCTGAGAGAAACCTTCAAAATG 1317
Qy 426 TAAACAATGTAAATTAAGCTTTTCACTGACCTTCAAAATTAAGCAAAATTAATTAACA 485
Db 1318 TAAAGAAATGTGTAAGACCTTTTAAACATCTTCAACCTTCAATTAAGATTAATTA 1377
Qy 486 TCATGTGATCTCTGCTTATCATATGATCATCTGCTGTTGTTTAAAAATTTCCAAAC 545
Db 1378 TACTGGAGAGAACCTTACAAATCTTAAAGATGTGAAAAAGCTTTTAAACCAATCTCA 1437
Qy 546 TTGCTAGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 605
Db 1438 ACTTACTGAACATTAAGAAATTCATCTGAGAGAAACCTTA---TGAATGTAAAAATG 1494
Qy 606 TGGTAAAGCTTGTGTTGGAAAAAGCTTATCTTCAATATGTTAAGTCATGATGA 662
Db 1495 TGGCAAGCTTTTACCAAGCTTCAATCTTACTAGACATTAAGAAAGTCATACAGA 1551

RESULT 4
US-09-949-016-4204
; Sequence 4204, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; PRIOR FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 4204
; LENGTH: 3798
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-4204

Query Match 7.3%; Score 90.8; DB 4; Length 3798;
Best Local Similarity 53.9%; Pred. No. 3.4e-11;
Matches 235; Conservative 0; Mismatches 192; Indels 9; Gaps 2;

Qy 69 AAGATATTTGACATATGAGGGTGTGATTAAGCTTATATGACCATCATTTATTAGA 128
Db 1848 AGAGAAACCTTACAAATGTGAAGATGTGCAAAAGCTTTAGCATCTTCAACCTTGC 1307
Qy 129 GCAACATTTAAGAACCCACAGTATGATGACCGTATTAATGTACAGTGAAGATTTGA 188
Db 1908 TAAACATTAAGAAATTCATCTGAGAGAAACCTTCAAAATGT-----GAAAGATGTG 1961
Qy 189 TAAAGATTTTGAAGAAATTCATCTGAAACACATATGTATTCATTCGGAATTA 248
Db 1962 CAAGAGCTTTTACCGCTTCTCAACCTTGTAAACATTAAGAAATTCATCTGAGAGAA 2021
Qy 249 ACCATTCATTTGTCAGTGTGTGTAAGGGGTTAAATTCGCAACAACCTTGAAGAGCA 308
Db 2022 ACCCTAACAAATGTAAAGATGTGCAAAAGCTTTTGAATTCCTCAACCTTGTATATCA 2081
Qy 309 TGAATATCCCATTAAGAAAGTCATTTAAATGTACATTTGAAATTTGCAAGAGCATTTTA 368
Db 2082 TAAAGATTAATCACTGAAGAGAAACCTTCAAAATGTAAAGATGTGACAAACCTTTTAA 2141
Qy 369 TAAACATCAATC---TTTAAGACATCATATATATCTGTTCAAGAAAAAATTAACGTG 425

Db 2142 GCGACTCTCAACCTTACTTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 2201
Qy 426 TAAACATGTAAATTAAGCTTTTCACTGACCTTAAATTAAGCAACATTAATTAACA 485
Db 2202 TGAAGATGTGCAAAAGCTTTTAAATGATCTTCAAAATCTTACTATTAATTAATTA 2261
Qy 486 TCATGTGATCTCT 501
Db 2262 TACTGGAGAGAAACCT 2277

RESULT 5
US-09-949-016-12227
; Sequence 12227, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; PRIOR FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 12227
; LENGTH: 156942
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)-(156942)
; OTHER INFORMATION: n = A,T,C or G
US-09-949-016-12227

Query Match 7.3%; Score 90.8; DB 4; Length 156942;
Best Local Similarity 53.9%; Pred. No. 1.1e-10;
Matches 235; Conservative 0; Mismatches 192; Indels 9; Gaps 2;

Qy 69 AAGATATTTGACATATGAGGGTGTGATTAAGCTTATATGACCATCATTTATTAGA 128
Db 152999 AGAGAAACCTTACAAATGTGAAGATGTGCAAAAGCTTTAGCATCTTCAACCTTGC 153058
Qy 129 GCAACATTTAAGAACCCACAGTATGATGACCGTATTAATGTACAGTGAAGATTTGA 188
Db 153059 TAAACATTAAGAAATTCATCTGAGAGAAACCTTCAAAATGT-----GAAAGATGTG 153112
Qy 189 TAAAGATTTTGAAGAAATTCATCTGAAACACATATGTATTCATTCGGAATTA 248
Db 153113 CAAGAGCTTTTACCGCTTCTCAACCTTGTAAACATTAAGAAATTCATCTGAGAGAA 153172
Qy 249 ACCATTCATTTGTCAGTGTGTGTAAGGGGTTAAATTCGCAACAACCTTGAAGAGCA 308
Db 153173 ACCCTAACAAATGTAAAGATGTGCAAAAGCTTTTGAATTCCTCAACCTTGTATATCA 153232
Qy 309 TGAATATCCCATTAAGAAAGTCATTTAAATGTACATTTGAAATTTGCAAGAGCATTTTA 368
Db 153233 TAAAGATTAATCACTGAAGAGAAACCTTCAAAATGTAAAGATGTGACAAACCTTTTAA 153292
Qy 369 TAAACATCAATC---TTTAAGACATCATATATATCTGTTCAAGAAAAAATTAACGTG 425
Db 153293 GCGACTCTCAACCTTACTTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 153352
Qy 426 TAAACATGTAAATTAAGCTTTTCACTGACCTTAAATTAAGCAACATTAATTAACA 485
Db 153353 TGAAGATGTGCAAAAGCTTTTAAATGATCTTCAAAATCTTACTATTAATTAATTA 153412

QY 486 TCATGTCGATCTCCT 501
| | | | |
Db 153413 TACTGGAGAGAAACCT 153428

RESULT 6

US-09-949-016-15946
Sequence 15946, Application US/09949016
Patent No. 6812339
GENERAL INFORMATION:
APPLICANT: VENTER, J. Craig et al.
TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
FILE REFERENCE: CL001307
CURRENT FILING DATE: 2000-04-14
PRIOR APPLICATION NUMBER: 60/241,755
PRIOR FILING DATE: 2000-10-20
PRIOR APPLICATION NUMBER: 60/237,768
PRIOR FILING DATE: 2000-10-03
PRIOR APPLICATION NUMBER: 60/231,498
PRIOR FILING DATE: 2000-09-08
NUMBER OF SEQ ID NOS: 207012
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 15946
LENGTH: 156950
TYPE: DNA
ORGANISM: Human
FEATURE:
NAME/KEY: misc_feature
LOCATION: (1)...(156950)
OTHER INFORMATION: n = A,T,C or G
US-09-949-016-15946

Query Match 7.3%; Score 90.8; DB 4; Length 156950;
Best Local Similarity 53.9%; Pred. No. 1.1e-10;

Matches 235; Conservative 0; Mismatches 192; Indels 9; Gaps 2;

QY 69 AAAGTATATTTGACATATGAGGGGTGATTAAGGCTTAATTCGACATCTATTATTAGA 128
Db 152999 AGAGAAACCTTACAAATGTGAGAAATGTGGCAAAAGCTTTTACCATCTTCAACCTTGC 153058
QY 129 GCAACATTTAAGAACCCACAGTAATGATGACCGTAAATGTAAGTACAGTGACGATTGTGA 188
Db 153059 TAAACATAGAGAAATCATCTGAGAGAAACCTTCAATGT-----GAAGAAATGTGG 153112
QY 189 TAAAGCATTTTTCAGAAATTCATTTGGAAACATATTGTATCATCTCCGAAAAAAA 248
Db 153113 CAAAGCTTTTACCGCTTCTTCAACCTTCTAAACATTAAGAAATTCATCTAGAGAGAA 153172
QY 249 ACCATTCCATTGTCAGTGTGTGTAAGGGGTAAATTCCTCGACAAACATTTGAAAAAGCA 308
Db 153173 ACCCTTAAGAAATGTGAGAAATGTGGCAAAAGCTTTTACCAATCTCTCAACCTTGTATATCA 153232
QY 309 TGAATTCACCCATACAAAGTCAATTAATGTACATTTGAAAAATGTCAAGAAACATTTTA 368
Db 153233 TAAAGATTAATCTAGTGAAGAGAAACCTTACAAATGTAAGATGTGACAAACCTTTTA 153292
QY 369 TAAACATCAATC---TTTAAAGCATATATTTATCTGTTTCATGAAAAAAACATTAACGTG 425
Db 153293 GGAACCTCAACCTTACTTAATCAATTAATCACTGTGAGAGAAACCTCTACAAATG 153352
QY 426 TAAACATGTAATTAAGTTTTCCTGACCTTCAAAATTAAGACACATTAATTAATAACA 485
Db 153353 TGAAGATGTGGCAAAAGCTTTTATGATCTTCAAAATCTTACTATACATTAAGTTTATCA 153412
QY 486 TCATGTCGATCTCCT 501
| | | | |
Db 153413 TACTGGAGAGAAACCT 153428

RESULT 7
US-09-949-016-485

Sequence 485, Application US/09949016
Patent No. 6812339
GENERAL INFORMATION:
APPLICANT: VENTER, J. Craig et al.
TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
FILE REFERENCE: CL001307
CURRENT FILING DATE: 2000-04-14
PRIOR APPLICATION NUMBER: 60/241,755
PRIOR FILING DATE: 2000-10-20
PRIOR APPLICATION NUMBER: 60/237,768
PRIOR FILING DATE: 2000-10-03
PRIOR APPLICATION NUMBER: 60/231,498
PRIOR FILING DATE: 2000-09-08
NUMBER OF SEQ ID NOS: 207012
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 485
LENGTH: 3839
TYPE: DNA
ORGANISM: Human
US-09-949-016-485

Query Match 7.2%; Score 89.2; DB 4; Length 3839;
Best Local Similarity 53.7%; Pred. No. 7.9e-11;
Matches 234; Conservative 0; Mismatches 193; Indels 9; Gaps 2;

QY 69 AAAGTATATTTGACATATGAGGGGTGATTAAGGCTTAATTCGACATCTATTATTAGA 128
Db 964 AGAGAAACCTTACAAATGTGAGAAATGTGGCAAAAGCTTTTACCATCTTCAACCTTGC 1023
QY 129 GCAACATTTAAGAACCCACAGTAATGATGACCGTAAATGTAAGTACAGTGACGATTGTGA 188
Db 1024 TAAACATAGAGAAATCATCTGAGAGAAACCTTCAATGT-----GAAGAAATGTGG 1077
QY 189 TAAAGCATTTTTCAGAAATTCATTTGGAAACATATTGTATCATCTCCGAAAAAAA 248
Db 1078 CAAAGCTTTTACCGCTTCTTCAACCTTCTAAACATTAAGAAATTCATCTAGAGAGAA 1137
QY 249 ACCATTCCATTGTCAGTGTGTGTAAGGGGTAAATTCCTCGACAAACATTTGAAAAAGCA 308
Db 1138 ACCCTTAAGAAATGTGAGAAATGTGGCAAAAGCTTTTACCAATCTCTCAACCTTGTATATCA 1197
QY 309 TGAATTCACCCATACAAAGTCAATTAATGTACATTTGAAAAATGTCAAGAAACATTTTA 368
Db 1198 TAAAGATTAATCTAGTGAAGAGAAACCTTACAAATGTAAGATGTGACAAACCTTTTA 1257
QY 369 TAAACATCAATC---TTTAAAGCATATATTTATCTGTTTCATGAAAAAAACATTAACGTG 425
Db 1258 GGAACCTCAACCTTACTTAATCAATTAATCACTGTGAGAGAAACCTCTACAAATG 1317
QY 426 TAAACATGTAATTAAGTTTTCCTGACCTTCAAAATTAAGACACATTAATTAATAACA 485
Db 1318 TGAAGATGTGGCAAAAGCTTTTATGATCTTCAAAATCTTACTATACATTAAGTTTATCA 1377
QY 486 TCATGTCGATCTCCT 501
| | | | |
Db 1378 TACTGGAGAGAAACCT 1393

RESULT 8

US-09-949-016-27359/c
Sequence 27359, Application US/09949016
Patent No. 6812339
GENERAL INFORMATION:
APPLICANT: VENTER, J. Craig et al.
TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
FILE REFERENCE: CL001307
CURRENT FILING DATE: 2000-04-14
PRIOR APPLICATION NUMBER: 60/241,755
PRIOR FILING DATE: 2000-10-20

;; PRIOR APPLICATION NUMBER: 60/237,768
;; PRIOR FILING DATE: 2000-10-03
;; PRIOR APPLICATION NUMBER: 60/231,498
;; PRIOR FILING DATE: 2000-09-08
;; NUMBER OF SEQ ID NOS: 207012
;; SOFTWARE: FastSeq for Windows Version 4.0
;; SEQ ID NO 27359
;; LENGTH: 601
;; TYPE: DNA
;; ORGANISM: Human
US-09-949-016-27359

Query Match
Best Local Similarity 51.4%; Pred. No. 1.6e-10;
Matches 252; Conservative 1; Mismatches 228; Indels 9; Gaps 2;

QY 7 GAAAGTGACGAAACCAATGATATCATCTTTAATATCTTCTTCTTCATCAGTCCC 66
DB 538 GAATGTGGCAAAAGCTTTTAAGCACTCCTCAGCCCTTCTAAACATTAATATATCATGCT 479
QY 67 AAAAGATATTTTGACATATGAAAGGTGATTAAGCTTAATGACACATCATTTATTA 126
DB 478 GGAAGAACTCTACAAATGTGAGAAATGTGGCAAAAGCTTTATCATCTTCAATCTT 419
QY 127 GAGCAACATTTTAAGAAACCAAGTATGATCGACCGTATTAATGTACAGTGAAGATTTG 186
DB 418 ACACACATTAAGATTAATTCATCTTAAGAGAAACCTTCCAGAGT-----GAAAGATGT 365
QY 187 GATTAAGCATTTTTCAGAAAATCATTTTGGAAAACATATTTGTATCATCTCCGAAAA 246
DB 364 GACAAACATTTATCTGTCTCTCAACCTTACTGAACATTAAGAAATTCATACAGAGAG 305
QY 247 AAACCATTCATTTTCAAGTGTGTGTGTAAGGGGTAAATCTCGACAAACCTTGAAGA 306
DB 304 AAAMCTCAAAATGTGAAATGTGGCAAAAGCATTTTGGCAAGCTTTCACCTTACTACA 245
QY 307 CATGAATCAACCCATACAAAGTCAATTAATGTACATTTTGAATTTGTCAAGAAAGCATTT 366
DB 244 CATAGAGATGACACCTGAGAGAAACCTTACAAATGTGAAGATGTGGCAAAAGCTTTT 185
QY 367 TATTAACATCAATC---TTTAAGACATCATATATATCTGTTCATGAAAAAACAATTACG 423
DB 184 AGCCAAATCTCAACCTTACTACACATTAAGATTAATCTGAGAGAAACCTTACAAA 125
QY 424 TGTAAACAATGTAATAAGTTTCACTCGACCTTCAAAATTAAGCAACAATTAATTA 483
DB 124 TGTGAAGAATGTGGCAAAAGCTTTTGAAGAAATCTTCACTTACTGAACATTAAGATTA 65
QY 484 CATCATGTG 493
DB 64 CATACTGGAG 55

RESULT 9
US-09-949-016-149638/c
;; Sequence 149638, Application US/09949016
;; Patent No. 6812339
;; GENERAL INFORMATION:
;; APPLICANT: VENTER, J. Craig et al.
;; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
;; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
;; FILE REFERENCE: CL001307
;; CURRENT APPLICATION NUMBER: US/09/949,016
;; PRIOR FILING DATE: 2000-04-14
;; PRIOR APPLICATION NUMBER: 60/241,755
;; PRIOR FILING DATE: 2000-10-20
;; PRIOR APPLICATION NUMBER: 60/237,768
;; PRIOR FILING DATE: 2000-10-03
;; PRIOR APPLICATION NUMBER: 60/231,498
;; PRIOR FILING DATE: 2000-09-08
;; NUMBER OF SEQ ID NOS: 207012
;; SOFTWARE: FastSeq for Windows Version 4.0
;; SEQ ID NO 149638

;; LENGTH: 601
;; TYPE: DNA
;; ORGANISM: Human
US-09-949-016-149638

Query Match
Best Local Similarity 51.4%; Pred. No. 1.6e-10;
Matches 252; Conservative 1; Mismatches 228; Indels 9; Gaps 2;

QY 7 GAAAGTGACGAAACCAATGATATCATCTTTAATATCTTCTTCTTCATCAGTCCC 66
DB 538 GAATGTGGCAAAAGCTTTTAAGCACTCCTCAGCCCTTCTAAACATTAATATATCATGCT 479
QY 67 AAAAGATATTTTGACATATGAAAGGTGATTAAGCTTAATGACACATCATTTATTA 126
DB 478 GGAAGAACTCTACAAATGTGAGAAATGTGGCAAAAGCTTTTATCATCTTCAATCTT 419
QY 127 GAGCAACATTTTAAGAAACCAAGTATGATCGACCGTATTAATGTACAGTGAAGATTTG 186
DB 418 ACACACATTAAGATTAATTCATCTTAAGAGAAACCTTCCAGAGT-----GAAAGATGT 365
QY 187 GATTAAGCATTTTTCAGAAAATCATTTTGGAAAACATATTTGTATCATCATTCGAAAA 246
DB 364 GACAAACATTTATCTGTCTCTCAACCTTACTGAACATTAAGAAATTCATACAGAGAG 305
QY 247 AAACCATTCATTTTCAAGTGTGTGTGTAAGGGGTAAATCTCGACAAACCTTGAAGA 306
DB 304 AAAMCTCAAAATGTGAAATGTGGCAAAAGCATTTTGGCAAGCTTTCACCTTACTACA 245
QY 307 CATGAATCAACCCATACAAAGTCAATTAATGTACATTTTGAATTTGTCAAGAAAGCATTT 366
DB 244 CATAGAGATGACACCTGAGAGAAACCTTACAAATGTGAAGATGTGGCAAAAGCTTTT 185
QY 367 TATTAACATCAATC---TTTAAGACATCATATATCTGTTCATGAAAAAACAATTACG 423
DB 184 AGCCAAATCTCAACCTTACTACACATTAAGATTAATCTGAGAGAAACCTTACAAA 125
QY 424 TGTAAACAATGTAATAAGTTTCACTCGACCTTCAAAATTAAGCAACAATTAATTA 483
DB 124 TGTGAAGAATGTGGCAAAAGCTTTTGAAGAAATCTTCACTTACTGAACATTAAGATTA 65
QY 484 CATCATGTG 493
DB 64 CATACTGGAG 55

RESULT 10
US-09-949-016-44799
;; Sequence 44799, Application US/09949016
;; Patent No. 6812339
;; GENERAL INFORMATION:
;; APPLICANT: VENTER, J. Craig et al.
;; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
;; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
;; FILE REFERENCE: CL001307
;; CURRENT APPLICATION NUMBER: US/09/949,016
;; PRIOR FILING DATE: 2000-04-14
;; PRIOR APPLICATION NUMBER: 60/241,755
;; PRIOR FILING DATE: 2000-10-20
;; PRIOR APPLICATION NUMBER: 60/237,768
;; PRIOR FILING DATE: 2000-10-03
;; PRIOR APPLICATION NUMBER: 60/231,498
;; PRIOR FILING DATE: 2000-09-08
;; NUMBER OF SEQ ID NOS: 207012
;; SOFTWARE: FastSeq for Windows Version 4.0
;; SEQ ID NO 44799
;; LENGTH: 601
;; TYPE: DNA
;; ORGANISM: Human
US-09-949-016-44799

Query Match
Best Local Similarity 54.4%; Pred. No. 2.1e-10;
Matches 252; Conservative 1; Mismatches 228; Indels 9; Gaps 2;

Matches 241; Conservative 0; Mismatches 193; Indels 9; Gaps 3;

QY 61 CGTCCAAAAGATATTTTGCACATATGAAGGTGTGATTAAGCCATATATGACCATCA 120
DB 56 CATCTAGAGAGAAACCTTACCAATGTGAAGATGTGGCAAGCATTTATATGCTCTCA 115
QY 121 TTATTAAGCAACATTTAAGAACCCACAGTAATGATGACCGTATTAATGTACAGTGAC 180
DB 116 ACCCTAATGACATATGAAGAGATGACACTGGAGAGAAACCTTACCAATGT-----GAA 169
QY 181 GATTGATTAAGCATTTTTCAGAAAATGACATTTGGAAAACATTTGTATCACTATCC 240
DB 170 GAATGTGCAAAATCTTTTACCAATCTTCAACCTTACTACATTAAGAAATTTCTACT 229
QY 241 GAAAAAACAATTCATTTTCACTGTGTGTGTAAGGGTTAATCTGACAACTTG 300
DB 230 GGAGAGAAACCTTACCAATGTGAAGATGTGGCAAGCCTTTTACGTGCTCAACTCT 289
QY 301 AAAAGCATGAATTCACCATACAAAGTCATTTAAATGTACATTGAAATTTGTCAAGAA 360
DB 290 ACTAAACATTAATATTAATCTGTAAGAGAAACCTTACCAATGTGAAAAATGTGGCAAA 349
QY 361 GCATTTTA-TAAACATCATCTTTA--AGACATCATATATTAATCTGTCATGAAAAAACA 417
DB 350 GCCTTTAAGCAGTCTTCAATCTTACTTAAACCATTAAGAAATTCATCTGAGAGAAACC 409
QY 418 TTAACGTGAACAAATGATTAAGTTTCACTGACCTTCAAAATTTAGCACAACTATAA 477
DB 410 TATTAATGTGAAGATGTGGCAAAATCTTTTAACCGGTCTTCAACTTTTACTAAACATAAG 469
QY 478 TTAACATCATGTGTGATCTCC 500
DB 470 GTAATTCATCTGAGATAAAC 492

RESULT 11
US-09-949-016-13027
; Sequence 13027, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; PRIOR FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 13027
; LENGTH: 45138
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(45138)
; OTHER INFORMATION: n = A,T,C or G
US-09-949-016-13027

Query Match 7.0%; Score 86.2; DB 4; Length 45138;
Best Local Similarity 54.4%; Pred. No. 8.4e-10;
Matches 241; Conservative 0; Mismatches 193; Indels 9; Gaps 3;

QY 61 CGTCCAAAAGATATTTTGCACATATGAAGGTGTGATTAAGCCATATATGACCATCA 120
DB 42121 CATCTAGAGAGAAACCTTACCAATGTGAAGATGTGGCAAGCATTTTATATGCTCTCA 42180
QY 121 TTATTAAGCAACATTTAAGAACCCACAGTAATGATGACCGTATTAATGTACAGTGAC 180

DB 42181 ACCCTAATGACATTAAGAGGATGCACTGGAGAGAAACCTTACCAATGT-----GAA 42234
QY 181 GATTGATTAAGCATTTTTCAGAAAATGACATTTGGAAAACATATGATATCACTATCC 240
DB 42235 GAATGTGCAAAATCTTTTACCAATCTTCAACCTTACTACATTAAGATTAATTTCTACT 42294
QY 241 GAAAAAACAATTCATTTTCACTGTGTGTGTAAGGGTTAATTTCTGACAACTTG 300
DB 42295 GGAGAGAAACCTTACCAATGTGAAGATGTGGCAAGCCTTTTACGTGCTCAACTCT 42354
QY 301 AAAAGCATGAATTCACCATACAAAGTCATTTAAATGTACATTGAAATTTGTCAAGAA 360
DB 42355 ACTAAACATTAATATTAATCTGTAAGAGAAACCTTCAAAATGTGAAAAATGTGGCAAA 42414
QY 361 GCATTTTA-TAAACATCATCTTTA--AGACATCATATATTAATCTGTCATGAAAAAACA 417
DB 42415 GCCTTTAAGCAGTCTTCAATCTTACTTAAACCATTAAGAAATTCATCTGAGAGAAACC 42474
QY 418 TTAACGTGAACAAATGATTAAGTTTCACTGACCTTCAAAATTTAGCACAACTATAA 477
DB 42475 TATTAATGTGAAGATGTGGCAAAATCTTTTAACCGGTCTTCAACTTTTACTAAACATAAG 42534
QY 478 TTAACATCATGTGTGATCTCC 500
DB 42535 GTAATTCATCTGAGATAAAC 42557

RESULT 12
US-09-949-016-27360/c
; Sequence 27360, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; PRIOR FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 27360
; LENGTH: 601
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-27360

Query Match 6.9%; Score 85.6; DB 4; Length 601;
Best Local Similarity 51.4%; Pred. No. 2.9e-10;
Matches 252; Conservative 0; Mismatches 229; Indels 9; Gaps 2;

QY 7 GAAAGTGAAGAAACCAATTCATATCTTTAATATCTTCTTCTCATCAGTCC 66
DB 499 GAATGTGCAAAAGCTTTTAAAGCATCTCTGACCTTGTCAAAATTAATATATCATGT 440
QY 67 AAAAGATATTTTGCACATATGAAGGTGTGATTAAGCCATATATGACCATATTA 126
DB 439 GGAGAGAAACCTTACCAATGTGAAGATGTGGCAAGCCTTTTAAATCATCTTCAATCT 380
QY 127 GAGCAATTTAAGAACCCACAGTAATGATGACCGTATTAATGTACAGTGACATTTGT 186
DB 379 ACGACATTAAGATATTAATCTTAAAGAGAAACCTTCAAGAGT-----GAAAGATGT 326
QY 187 GATTAAGATTTTTCAGAAAATGACATTTGAAAACATATTTGTATCATCTCGAAAAA 246
DB 335 GACAAAGCATTTATCTGTGCTCTCACCCTTACTGAACATTAAGAGATTCATACAGAGG 266

QY 247 AACCATTCATGTTGAGTGTGTGTAAAGGGTTAATCTCGACAACCTTGAAAAA 306
DB AAAACCTTAACAATGTGAAGATGTGGCAAGCATTTAGCCAGCCTTGACACCTTACTACA 206
QY 307 CATGAATCAACCCATACAAAGTCAATTAATGTACATTTGAAATTTGTCAAGAACCTTT 366
DB 205 CATTAAGGATGTGCACTGTGAGAGAAACCTTACAAATGTGAAGATGTGCAAGCTTTT 146
QY 367 TATTAACATCAATC---TTTAAGACATCATATATTATCTGTTCATGAAAAACATTAACG 423
DB 145 AGCCAACTCTCAACCCCTTACTACACATTAAGATATTCTACTGAGAGAAACCTTACAA 86
QY 424 TGTAAACATGTATAAAGTTTCTGACCTTCAAAATTAAGCAACAATAATTAAAA 483
DB 85 TGTGAAGATGTGGCAAGCTTTTGAAGAAATCTTCACTCTTACTGAACATTAAGATATT 26
QY 484 CATCATGTGTG 493
DB 25 CATACTGGAG 16

RESULT 13

US-09-949-016-149637/c
; Sequence 149637, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; PRIOR FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FASTSEQ for Windows Version 4.0
; SEQ ID NO 149637
; LENGTH: 601
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-149637

Query Match 6.9%; Score 85.6; DB 4; Length 601;

Best Local Similarity 51.4%; Pred. No. 2.9e-10;
Matches 252; Conservative 0; Mismatches 229; Indels 9; Gaps 2;

QY 7 GAAAGTGAAGAAACCAATGATATCATCTTTAATATCTTCTTCTTATGACGTC 66
DB 499 GAATGTGGCAAGCTTTTAAGACCTCTCGAGCCCTTGCTAAACATAAATATATACATGT 440
QY 67 AAAAGTATATTTGACATATGAAGGTGTGATTAAGCTATATGACACCATATATTA 126
DB 439 GGAAGAAATCTTAACAATGTGAGATGTGGCAAGCTTTTANTCAATTTCTTAAATCTT 380
QY 127 GAGCAACATTTAAGAACCAAGTATGATGACCCGTATTAATGTATGACGTGAGCATTTG 186
DB 379 ACACACATTAAGATTAATTTACTTAAGAGAAACCTTCCAAGAGT-----GAAAGATGT 326
QY 187 GATTAAGCATTTTTCAGAAAATCACATTTGAAAAACATATTTGTATACATTTCCGAAAA 246
DB 325 GACAAAGCATTTTATGTGTCTCTCARCCCTTAAGTGAACATTAAGAAATTCATACAGAGAG 266
QY 247 AAAACATTTTCATGTTGAGTGTGTGTAAAGGGTTAATCTCGACAACCTTGAAAAA 306
DB 265 AAAACCTTAACAATGTGAAGATGTGGCAAGCATTTAGCCAGCCTTGACACCTTACTACA 206
QY 307 CATGAATCAACCCATACAAAGTCAATTAATGTACATTTGAAATTTGTCAAGAACCTTT 366
DB 205 CATTAAGGATGTGCACTGTGAGAGAAACCTTACAAATGTGAAGATGTGCAAGCTTTT 146

QY 367 TATTAACATCAATC---TTTAAGACATCATATATTATCTGTGATGAAAAACATTAACG 423
DB 145 AGCCAACTCTCAACCCCTTACTACACATTAAGATATTCTACTGAGAGAAACCTTACAA 86
QY 424 TGTAAACATGTATAAAGTTTCTGACCTTCAAAATTAAGCAACAATAATTAAAA 483
DB 85 TGTGAAGATGTGGCAAGCTTTTGAAGAAATCTTCACTCTTACTGAACATTAAGATATT 26
QY 484 CATCATGTGTG 493
DB 25 CATACTGGAG 16

RESULT 14

US-09-949-016-27361/c
; Sequence 27361, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; PRIOR FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FASTSEQ for Windows Version 4.0
; SEQ ID NO 27361
; LENGTH: 601
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-27361

Query Match 6.7%; Score 83.2; DB 4; Length 601;

Best Local Similarity 52.5%; Pred. No. 1e-09;
Matches 231; Conservative 1; Mismatches 199; Indels 9; Gaps 2;

QY 57 ATCACTTCCCAAAAGATATTTTGCACATTAAGAGGTGTGATTAAGCCCTATATGACC 116
DB 554 AACTCATACTGAAGAGAAACCTTACAAATGTAAAGATGTGACAAACCTTTTAAGCAGCT 495
QY 117 ATCATTTATGAGCAACATTTAAGAACCCACAGTATGATGACCGTATTAATGTACAGT 176
DB 494 CTCGAACCTTACTTAACAATTAATATACATGCTGAGAGAAACCTTCAAAATGT----- 440
QY 177 GAGCATGTGTATTAAGCATTTTTCAGAAAATCAATTTGAAACATATTTGTATCACA 236
DB 439 -GAAGATGTGGCAAGCTTTTATGATGATCTTCAAACTTACTATACATTAAGTTTATTTCA 381
QY 237 TTCCGAAAAAAACCATTCATTTGTCAGTGTGTGTGTAAGGGGTATTTCTGACACACA 296
DB 380 TACTGAGAGAAACCTTACAAAGTGAAGATGTGCAAGCATTTAACTGTGCTCTCAAG 321
QY 297 CTGAAAAGACATGAATATCCCATCAAAAGTCAATTTAATGTATGACATTTGAAAAATTGTCA 356
DB 320 CCTTACTTAACATTAAGAAATTTCTATCTAGAGAGAAACCTTCAAAATGTAAAGATGTGG 261
QY 357 AGAAGCATTTTATTAACATCAATCTTTA--AGACATCATATATTTATCTGTGATGAAAA 413
DB 260 CAAGCATTTATATGTGCTTCAACCTTAACCTAGACATTAAGAGATACACCTGAGAGAA 201
QY 414 AACATTAAGTGTAAACATGTATTAAGTTTTCATCGACCTTCAAAATTTAGACACACA 473
DB 200 GCCCTTACAAATGTGAAGATGTGGCAAGCTTTTACCGCTTCTCAACCTTACTAAGCA 141
QY 474 TAAATTTAAACATCATGTGTG 493

Db 140 TAAAGCAATTCATCTGAG 121

RESULT 15

US-09-949-016-149636/c

/ Sequence 149636, Application US/09949016

/ Patent No. 6812339

/ GENERAL INFORMATION:

/ APPLICANT: VENTER, J. Craig et al.

/ TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED

/ WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF

/ FILE REFERENCE: CL001307

/ CURRENT APPLICATION NUMBER: US/09/949,016

/ PRIORITY FILING DATE: 2000-04-14

/ PRIOR APPLICATION NUMBER: 60/241,755

/ PRIOR FILING DATE: 2000-10-20

/ PRIOR APPLICATION NUMBER: 60/237,768

/ PRIOR FILING DATE: 2000-10-03

/ PRIOR APPLICATION NUMBER: 60/231,498

/ NUMBER OF SEQ ID NOS: 207012

/ SOFTWARE: FastSeq for Windows Version 4.0

/ SEQ ID NO 149636

/ LENGTH: 601

/ TYPE: DNA

/ ORGANISM: Human

US-09-949-016-149636

Query Match 6.7%; Score 83.2; DB 4; Length 601;

Best Local Similarity 52.5%; Pred. No. 1e-09;

Matches 231; Conservative 1; Mismatches 199; Indels 9; Gaps 2;

QY 57 ATACGCTCCCAAAAAGTATATTTGCACATATGAGGGTGTGATTAAGCTATATGACC 116
DB 554 AACTGCTACTGAGAGAAACCTCAAAATGTAAGATGTGACAAAACCTTTAAGGACT 495
QY 117 ATCATTATTAGAGCAACATTTAAGAACCAAGTAATGATCGACCGTATTAATGACGT 176
DB 494 CTCACCTTACTTAACATTAATAATATATGATGCTGAGAGAAACCTCAAAATGT----- 440
QY 177 GGACGATTGTGATTAAGCATTTTTCAGAAATATCATTTGGAACATATTTGTATCACA 236
DB 439 -GAAGATGTGGCAAGCTTTTAATGATCTTCAAACTTACTATACATAAAGTTTATTC 381
QY 237 TTCGAAAAAAAACATTCATTTGTCAGTGTGTAAGGGGTTAATTTCTGACACACA 296
DB 380 TACTGAGAGAAACCTTACAGTGTGAGAAATGTGCAAAAGCATTTAAGTCTCTCAG 321
QY 297 CTTGAAAAGACATGAATCACCCCATACAAAGTCATTTAATGATGACATTTGAAAATTTGCA 356
DB 320 CCTTACTTAACATTAAGAAATTCATCTAGAGAGAAACCTTCAAAATGTAAGAAATGTG 261
QY 357 AGAAGCATTTTATTAACATCATCTTTA--AGACATCATATATATCTGTTCATGAAA 413
DB 260 CAAGCATTTTAATGTCTTCAACCTTACATGACATTAAGAGATACACATGAGAGAA 201
QY 414 AACATTACGTGTAACATGTAATTAAGTTTCACTCGACCTTCAAAATTAAGACACACA 473
DB 200 GCCCTTACAAATGTGAAGATGTGGCAAGCTTTTACCGTTCTCAACCTTACTAAGCA 141
QY 474 TAAATTAAACATCATGCTG 493
DB 140 TAAAGCAATTCATCTGAG 121

Search completed: February 9, 2005, 07:51:09
Job time : 227.728 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using bw model

Run on: February 9, 2005, 05:36:40 ; Search time 2154.88 Seconds
(without alignments)
3301.382 Million cell updates/sec

Title: US-09-831-804-1_COPY_720_1955
Perfect score: 1236
Sequence: 1 atgagtggaagagtcagcaaac.....aaacatcagtgattctcga 1236

Scoring table:
IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 4313806 seqs, 287871033 residues

Total number of hits satisfying chosen parameters: 8627612

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

Published Applications NA:*

- 1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq:*
- 2: /cgn2_6/ptodata/1/pubpna/PCT_NEW_PUB.seq:*
- 3: /cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq:*
- 4: /cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq:*
- 5: /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq:*
- 6: /cgn2_6/ptodata/1/pubpna/PCTUS_PUBCOMB.seq:*
- 7: /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq:*
- 8: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq:*
- 9: /cgn2_6/ptodata/1/pubpna/US09A_PUBCOMB.seq:*
- 10: /cgn2_6/ptodata/1/pubpna/US09B_PUBCOMB.seq:*
- 11: /cgn2_6/ptodata/1/pubpna/US09C_PUBCOMB.seq:*
- 12: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq:*
- 13: /cgn2_6/ptodata/1/pubpna/US10_PUBCOMB.seq:*
- 14: /cgn2_6/ptodata/1/pubpna/US10B_PUBCOMB.seq:*
- 15: /cgn2_6/ptodata/1/pubpna/US10C_PUBCOMB.seq:*
- 16: /cgn2_6/ptodata/1/pubpna/US10E_PUBCOMB.seq:*
- 17: /cgn2_6/ptodata/1/pubpna/US10F_PUBCOMB.seq:*
- 18: /cgn2_6/ptodata/1/pubpna/US10F_PUBCOMB.seq:*
- 19: /cgn2_6/ptodata/1/pubpna/US11_NEW_PUB.seq:*
- 20: /cgn2_6/ptodata/1/pubpna/US11_NEW_PUB.seq:*
- 21: /cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq:*
- 22: /cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1232.8	99.7	1239	US-10-032-585-6493	Sequence 6493, Ap
2	99.4	8.0	2132	US-10-094-749-726	Sequence 726, App
3	96.6	7.8	2320	US-10-029-386-20763	Sequence 1054, Ap
4	92	7.4	994	US-10-029-386-20763	Sequence 20763, A
5	92	7.4	2873	US-10-755-889-369	Sequence 369, App
6	91.8	7.4	2110	US-10-104-047-1778	Sequence 1778, Ap
7	91.8	7.4	3078	US-10-104-047-6222	Sequence 622, App
8	90.4	7.3	2114	US-10-094-749-692	Sequence 692, App
9	90.2	7.3	2230	US-10-094-749-499	Sequence 499, App
10	89.8	7.3	590	US-10-029-386-13298	Sequence 13298, A
11	89.2	7.2	2221	US-10-108-260A-29	Sequence 29, Appl

12	89.2	7.2	3839	US-10-172-118-910	Sequence 910, App
13	89.2	7.2	3839	US-10-342-887-910	Sequence 910, App
14	89	7.2	3502	US-10-098-841-48	Sequence 48, Appl
15	88.6	7.2	1466	US-10-029-386-25074	Sequence 25074, A
16	88.4	7.2	1410	US-10-029-386-20532	Sequence 20532, A
17	88.4	7.2	2509	US-10-381-372-11	Sequence 11, Appl
18	88.4	7.2	4132	US-10-198-846-13133	Sequence 13133, A
19	87.6	7.1	583	US-10-029-386-11093	Sequence 13093, A
20	87.6	7.1	1757	US-10-203-052B-6	Sequence 6, Appl
21	87.2	7.1	831	US-10-264-049-975	Sequence 975, Appl
22	86.8	7.0	2662	US-10-363-616-66	Sequence 66, Appl
23	86.2	7.0	1911	US-10-029-386-22636	Sequence 22636, A
24	86.2	7.0	1422	US-10-029-386-20322	Sequence 20322, A
25	86.2	7.0	1725	US-10-029-386-20556	Sequence 20556, A
26	86.2	7.0	1277	US-10-029-386-20367	Sequence 20367, A
27	86.2	7.0	2558	US-09-764-864-1664	Sequence 1664, Ap
28	86.2	7.0	2558	US-09-764-864-1665	Sequence 1665, Ap
29	86	7.0	2664	US-10-108-260A-511	Sequence 511, App
30	86	7.0	2690	US-10-723-860-8039	Sequence 8039, Ap
31	86	7.0	3400	US-10-104-047-1037	Sequence 1037, Ap
32	85.4	6.9	2505	US-10-108-260A-808	Sequence 808, App
33	84.8	6.9	2239	US-10-094-749-1084	Sequence 1084, Ap
34	84.4	6.8	1191	US-10-029-386-22992	Sequence 22992, A
35	84.4	6.8	3138	US-10-104-047-83	Sequence 83, Appl
36	84.2	6.8	1578	US-10-029-386-20214	Sequence 20214, A
37	84	6.8	565	US-10-029-386-3970	Sequence 3970, Ap
38	84	6.8	1619	US-10-422-522-46	Sequence 46, Appl
39	84	6.8	3617	US-10-029-386-22868	Sequence 22868, A
40	84	6.8	5820	US-10-363-616-44	Sequence 44, Appl
41	83.6	6.8	596	US-10-029-386-2664	Sequence 6664, Ap
42	83.2	6.7	1101	US-10-029-386-25405	Sequence 25405, A
43	83.2	6.7	2241	US-10-108-260A-1289	Sequence 1289, Ap
44	83.2	6.7	2525	US-10-108-260A-1358	Sequence 1358, Ap
45	83	6.7	500	US-10-029-386-6818	Sequence 6818, Ap

ALIGNMENTS

RESULT 1					
US-10-032-585-6493					
; Sequence 6493, Application US/10032585					
; Publication No. US20030180953A1					
; GENERAL INFORMATION:					
; APPLICANT: Terry, Roemer D.					
; APPLICANT: Bo, Jlang					
; APPLICANT: Charles, Boone					
; APPLICANT: Howard, Bussey					
; TITLE OF INVENTION: Gene Disruption Methodologies for Drug Target Discovery					
; FILE REFERENCE: 10182-005-999					
; CURRENT APPLICATION NUMBER: US/10/032,585					
; CURRENT FILING DATE: 2001-12-20					
; NUMBER OF SEQ ID NOS: 8000					
; SOFTWARE: PatentIn version 3.1					
; SEQ ID NO 6493					
; LENGTH: 1239					
; TYPE: DNA					
; ORGANISM: Candida albicans					
US-10-032-585-6493					
Query Match					
Best Local Similarity 99.7%; Score 1232.8; DB 16; Length 1239;					
Matches 1234; Conservative 0; Mismatches 2; Indels 0; Gaps 0;					
QY	1	ATGAGTGAAGAGTGAAGAACCAATGCATCTTAAATATCTTCTTCATCA	60		
DB	1	ATGAGTGAAGAGTGAAGAACCAATGCATCTTAAATATCTTCTTCATCA	60		
QY	61	CGTCCCAAAAGTATATTTGCAATATGAAGGTGTCATTAACCTTATCGACATCA	120		
DB	61	CGTCCCAAAAGTATATTTGCAATATGAAGGTGTCATTAACCTTATCGACATCA	120		
QY	121	TTATTGAGCAACATTAAAGACCAAGTAATGATCGACCGTATAAATGACAGTGAC	180		

Db 121 TTATGAGACCAATTTAAAGAACCCACAGTAATGACCGGTAAATGTAACAGGAG 180
QY 181 GATTGATGAAAGCATTTTTCAGAAATCAATTTGAAACACATATTGATCAATTC 240
Db 181 GATTGATGAAAGCATTTTTCAGAAATCAATTTGAAACACATATTGATCAATTC 240
QY 241 GAAAAAAACCATTCATTTGTCAGTGTGTGTAAGGGGTTAATTCGACAACTTG 300
Db 241 GAAAAAAACCATTCATTTGTCAGTGTGTGTAAGGGGTTAATTCGACAACTTG 300
QY 301 AAAAGCATGAATCAACCATCAAGTCAATTAATGTAATGTAATGTAATGTAATG 360
Db 301 AAAAGCATGAATCAACCATCAAGTCAATTAATGTAATGTAATGTAATGTAATG 360
QY 361 GATTGATGAAAGCATTTTTCAGAAATCAATTTGAAACACATATTGATCAATTC 420
Db 361 GATTGATGAAAGCATTTTTCAGAAATCAATTTGAAACACATATTGATCAATTC 420
QY 421 AGGTGTAACATGTAATTAAGTTTCACTGACCTTCAAAATTTAGCAACATTAAT 480
Db 421 AGGTGTAACATGTAATTAAGTTTCACTGACCTTCAAAATTTAGCAACATTAAT 480
QY 481 AAAACATCATGTCGATCTCTGCTTATCATGTCATCTCTGCTTATTAATAATTC 540
Db 481 AAAACATCATGTCGATCTCTGCTTATCATGTCATCTCTGCTTATTAATAATTC 540
QY 541 CAAACTTGGTCAGTATTTACATTTTCAATTAACATCTGATCCAAACCTTAATGCT 600
Db 541 CAAACTTGGTCAGTATTTACATTTTCAATTAACATCTGATCCAAACCTTAATGCT 600
QY 541 CAAACTTGGTCAGTATTTACATTTTCAATTAACATCTGATCCAAACCTTAATGCT 600
Db 541 CAAACTTGGTCAGTATTTACATTTTCAATTAACATCTGATCCAAACCTTAATGCT 600
QY 601 AAATGTGTAAAGTGTGTGTGGAAGAAAGGTTTACTTCAATTAATGTAATGTAAT 660
Db 601 AAATGTGTAAAGTGTGTGTGGAAGAAAGGTTTACTTCAATTAATGTAATGTAAT 660
QY 661 GATTCTACATGATCAAAATTAAGTGTGATTAATGTAATGTAATGTAATGTAAT 720
Db 661 GATTCTACATGATCAAAATTAAGTGTGATTAATGTAATGTAATGTAATGTAAT 720
QY 721 AAAAATGATTAATGTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 780
Db 721 AAAAATGATTAATGTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 780
QY 781 TTAAGGAAATCTGAAGTGAAGAAATTAAGAACTTAATGATCAAGATCGAAATTAAT 840
Db 781 TTAAGGAAATCTGAAGTGAAGAAATTAAGAACTTAATGATCAAGATCGAAATTAAT 840
QY 841 AATTGTCATGATTAAGAAACAGAAATTAAGTGAAGAAATGAAGATGAAGAA 900
Db 841 AATTGTCATGATTAAGAAACAGAAATTAAGTGAAGAAATGAAGATGAAGAA 900
QY 901 GATTGCTGATGAAGAAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 960
Db 901 GATTGCTGATGAAGAAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 960
QY 961 ATTAATCAATTTACTGCTTCTTGAAGGTTCAAGAGTGTCTTAATCTTAATCTGAAT 1020
Db 961 ATTAATCAATTTACTGCTTCTTGAAGGTTCAAGAGTGTCTTAATCTTAATCTGAAT 1020
QY 1021 AGTGGAGAAAGTCAATTTGCTTAAGAAATTAATGATGAATGTTTCTAGAGAAAT 1080
Db 1021 AGTGGAGAAAGTCAATTTGCTTAAGAAATTAATGATGAATGTTTCTAGAGAAAT 1080
QY 1081 GATTGATGTCGATTTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 1140
Db 1081 GATTGATGTCGATTTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 1140
QY 1141 AATAGATAGAAAAAGAAAGAACTCCAGAAAGTGAAGTGAAGTGAAGTGAAGTGAAG 1200
Db 1141 AATAGATAGAAAAAGAAAGAACTCCAGAAAGTGAAGTGAAGTGAAGTGAAGTGAAG 1200
QY 1201 GATTGATGTCGATTTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 1260
Db 1201 GATTGATGTCGATTTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 1260

Db 1201 GATTGATGTCGATTTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 1236
RESULT 2
US-10-094-749-726
Sequence 726, Application US/10094749
GENERAL INFORMATION:
APPLICANT: ISOGAI, TAKAO
APPLICANT: SUGIYAMA, TOMOYASU
APPLICANT: OTSUKI, TETSUJI
APPLICANT: WAKAMATSU, AI
APPLICANT: SATO, HIROYUKI
APPLICANT: ISHII, SHIZUKO
APPLICANT: YAMAMOTO, JUN-ICHI
APPLICANT: ISONO, YUUKO
APPLICANT: HIO, YURI
APPLICANT: OTSUKA, KAORU
APPLICANT: NAGAI, KEIICHI
APPLICANT: IRIE, RYOTARO
APPLICANT: TAMECHIKA, ICHIRO
APPLICANT: SEKI, NAOHICO
APPLICANT: YOSHIKAMA, TSUTOMU
APPLICANT: OTSUKA, MOTOTYUKI
APPLICANT: NAGAHARI, KENJI
APPLICANT: MASUHO, YASUHIKO
TITLE OF INVENTION: NOVEL, FULL-LENGTH cDNA
FILE REFERENCE: 08435/0160
CURRENT APPLICATION NUMBER: US/10/094,749
CURRENT FILING DATE: 2002-03-12
PRIOR APPLICATION NUMBER: 60/350,435
PRIOR FILING DATE: 2002-01-24
PRIOR APPLICATION NUMBER: JP 2001-328381
PRIOR FILING DATE: 2001-09-14
NUMBER OF SEQ ID NOS: 3381
SOFTWARE: Patentin Ver. 2.1
SEQ ID NO 726
LENGTH: 2132
TYPE: DNA
ORGANISM: Homo sapiens
US-10-094-749-726
Query Match 8.0%; Score 99.4; DB 17; Length 2132;
Best Local Similarity 50.6%; Pred. No. 4e-09;
Matches 297; Conservative 0; Mismatches 281; Indels 9; Gaps 2;
QY 7 GAAAGTGAAGAAACCAATTCATATCATTTTAATCTTCTTCTTCTTCAACAGTCCC 66
Db 1384 GATGTGGCAAGCTTTTAAACCAAGTCTCAATCTTACTACATPAAGAAATTCATACT 1443
QY 67 AAAAGTATATTGGACATATGTAAGGGTGTGTAAGGCTTAATGACCATCATTTATTA 126
Db 1444 GGAGGAAATCTTACAAATGTAAGAAATGTGGCAAGCTTTTATGATCTTCAAACTT 1503
QY 127 GAGCAACATTTAAGAACCCACAGTAATGATGACCGGTAAATGTAAGTGAAGTGAATGT 186
Db 1504 ACTGAACATPAAGAAATTCATCTGAGAGAAACCTTACACATGT-----GAAAGATGT 1557
QY 187 GATTAAGCATTTTTCAGAAATCAATTTGGAACACATATTGTAATCATTTCCGAAAAA 246
Db 1558 GGCAAGGCTTTTAAACCATCTCCACACCTGTGTACACATPAAGTATTCATCTGAGAG 1617
QY 247 AAACCATTCATTTGTCAGTGTGTGTAAGGGGTTAATTCGCAACACATTTGAAAAA 306
Db 1618 AAACCTTCAATGTAAGAAATGTGTAAAGCTTTAACCACTCTCACACCTTACTAGA 1677
QY 307 CATGAATCACCCATACAAAGTCAATTTAATGTAATGTAATGTAATGTAATGTAATGTAAT 366
Db 1678 CATTAAGAAATTCATACGAGAGAGAAACCTTCAATGTAATGTAATGTAATGTAATGTAAT 1737
QY 367 TATTAACATCAATCTTTAA--GACATCATATATTATCTGTTCATGAAAAAACATTAAC 423
Db 1738 AACGAGTCTCAACCTTACTGAGATTAAGAAATTCATATCTGTGAGAAACCTTACAAA 1797

RESULT 4
 US-10-029-386-20763/c
 ; Sequence 20763, Application US/10029386
 ; Publication No. US20030194704A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Penn, Sharon G.
 ; APPLICANT: Hank, David R.
 ; APPLICANT: Hanzel, David K.
 ; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
 ; TITLE OF INVENTION: EXPRESSION ANALYSIS TWO
 ; FILE REFERENCE: AROMICA-X-2
 ; CURRENT APPLICATION NUMBER: US/10/029,386
 ; CURRENT FILING DATE: 2001-12-20
 ; NUMBER OF SEQ ID NOS: 34288
 ; SOFTWARE: Anomax Sequence Listing Engine vers. 1.1
 ; SEQ ID NO 20763
 ; LENGTH: 994
 ; TYPE: DNA
 ; ORGANISM: Homo sapiens
 ; FEATURE:
 ; OTHER INFORMATION: MAP TO AC010620.3
 ; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 2.4
 ; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 3.1
 ; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 3
 ; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 4.1
 ; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 2.6
 ; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 3.4
 ; OTHER INFORMATION: NT HIT: g413650821, EVALUE 0.00e+00
 ; OTHER INFORMATION: EST HUMAN HIT: B1093435.1, EVALUE 0.00e+00
 ; OTHER INFORMATION: SWISSPROT HIT: Q03923, EVALUE 0.00e+00
 US-10-029-386-20763
 Query Match 7.4%; Score 92; DB 16; Length 994;
 Best Local Similarity 52.0%; Pred. No. 7.7e-08;
 Matches 259; Conservative 0; Mismatches 230; Indels 9; Gaps 2;

Oy	424	TGTAACATGTAAATTAACTTTTCACTGCAGCCTTGAAAATTACACACAATAAATPAAA	483
Dd	349	TGTGAAGAATGTGGCAAAGCCCTTTAACCACTCCTCAAACTACTTAAGATTAAT	250
Oy	484	CATCATGTGGATCTTCCT	501
Dd	289	CATACCTGGAGAGAAACCT	272

```

RESULT 5
US-10-755-889-369
; Sequence 369, Application US/10755889
; Publication No. US20040171823A1
; GENERAL INFORMATION:
; APPLICANT: Bristol-Myers Squibb Company
; TITLE OF INVENTION: POLYNUCLEOTIDES AND POLYPEPTIDES ASSOCIATED WITH THE NF-KB
; TITLE OF INVENTION: PATHWAY
; FILE REFERENCE: D0284 NP
; CURRENT APPLICATION NUMBER: US/10/755,889
; CURRENT FILING DATE: 2004-01-13
; PRIOR APPLICATION NUMBER: U.S. 60/440,068
; PRIOR FILING DATE: 2003-01-14
; PRIOR APPLICATION NUMBER: U.S. 60/469,757
; PRIOR FILING DATE: 2003-05-12
; NUMBER OF SEQ ID NOS: 823
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 369
; LENGTH: 2873
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-755-889-369

```

Query Match	7.4%	Score 92;	DB 18;	Length 2873;
Best Local Similarity	52.2%;	Pred. No. 1.2e-07;		
Matches 256;	Conservative 0;	Mismatches 225;	Indels 9;	Gaps 2;

Oy	7	GAAAGTGCAGAAACCAAAATGCATATCATCTTTAAATCTCTCTCTTCAACAGTCC	66
Db	1621	GAATGTCCCAAAACATTTTGCATGCTTTCAACCTTACTCAACATATAAAGATCCAAACT	1688
Oy	67	AAAAAGTATATTGGACATATGAAAGGTGTGATTAAGCTTATATCGAACATCATTATTA	126
Db	1681	AGAGTGAATTTCTCAAAATGTGAACCATATGGAAAGGCTTTTAACTGGTCTCCAACCTT	1740
Oy	127	GAGCAACATTTAAAGAACCCACAGTATATGATCGACCGTATTAATGTACAGTGAACGATGT	186
Db	1741	AATAAACATTAGAGAAATTCATCTGGAAGAAAACTTCAAAATGTAA-----GAATGT	1798
Oy	187	GATAAAGCATTTTCAGAAATATCATATTGGAAACAATTTGTATATCATTTCCGAAAAA	246
Db	1795	GGCAAAGCCTTTAACAGACCTCAACCTTATTAAGACATTAAGAGAAATTCATCTGAAG	1854
Oy	247	AAACCATTCATTTGTTAGTGTGTGTAAAGGGGTTAATTCGCAACACCTGTAAGAA	306
Db	1855	AAACCTTCMAATGTGAAGAAATGTGGCAAGCCTTTAACAGTATCGACCTTTCTTCA	1911
Oy	307	CATGAATTCACCCATACAAAGTCATTTAAATGTACATTTGAAATAATGTCAAGAACATTT	366
Db	1915	CATATATATATTCATACTGGGGAAATTCCTTCAATATGTGAAGAAATGTGTAGACTTTT	1974
Oy	367	TATTAACATCATCTTTAAGACATATATATATATCTGT---CATGAAAAAACAATTAA	423
Db	1975	AAACAGCCTCAAGCCTTACGAACATTAAGTTATTCATACCGAGAGAAACGTTATGA	2034
Oy	424	TGTAAACATGTAAATTAAGTTTTCACCTGACCTTCMAAATTAAGACACATTAATTA	483
Db	2035	TGTGAAGATGGCGCAAGCCTTTTAACGATCTCTAAAACCTTACGAAACATTAAGTCA	2094
Oy	484	CATCATGGTG 493	
Db	2095	CATACCTGGAG 2104	

```

RESULT 6
US-10-104-047-1778
; Sequence 1778, Application US/10104047
; Publication No. US20030236392A1
; GENERAL INFORMATION:
; APPLICANT: HELIX RESEARCH INSTITUTE
; TITLE OF INVENTION: NO. US20030236392A1 full length cDNA
; FILE REFERENCE: H1-A0105
; CURRENT APPLICATION NUMBER: US/10/104,047
; CURRENT FILING DATE: 2002-03-25
; PRIOR APPLICATION NUMBER:
; PRIOR FILING DATE:
; NUMBER OF SEQ ID NOS: 4096
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1778
; LENGTH: 2110
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-104-047-1778

```

Query Match	7.4%;	Score 91.8;	DB 17;	Length 2110;
Best Local Similarity	51.7%;	Pred. No. 1.1e-07;		
Matches 314;	Conservative 0;	Mismatches 277;	Indels 16;	Gaps 4;

QY	6	CGTCCAAAAGATGATATTTGGACATATGAAAGGGGTGATAAAGGCTATTAATCGACATCA	120
Db	1253	CATACCGAAGGAAACCTTACAAATGTAAAGATGTGCAAAAGCTTTTAAACACTCTCA	1312
QY	121	TTATTAAGCAACAATTAAAGAACCCACAGTAATGTCACCGTAAATGTACAGTGC	180
Db	1313	GCCCTTACTACACATTAAGAGAAATTCACACTGAGAGAAACCTTCAAAATGT-----GAA	1366
QY	181	GATTGTGATAAAGCATTTTTCAGAAAATCACATTTGGAAAACAATTTGTATCATCATTC	240
Db	1367	GAATGTGCAAAAGCTTTTAAACCGATCTCTCAAAACTTACTGAAACATTAAGAACTTCATCT	1428
QY	241	GAATAAAAAACCATTCACATGTTCAGTGTGTGTAAAGGGGTAAATTCGCAACAACCTTG	300
Db	1427	GGAAAGAAACCTTCAAAATGTGAAGATGTGCAAAAGCTTTATCCAACTCTCAAAACTT	1488
QY	301	AAAAGACATGAATCACCCATACAAAGTCATTTAAATGTACATTTGAAATTTGCAGAA	360
Db	1487	ACTGAACATTAATAAAAAATTCATTCGAGAGATACCTTACAAAGTGTGAAGATGTGGCAAA	1548
QY	361	GCATTTTATTAACATCAATCTTTA---AGACATCATATATATCGTTCATGAAAAACA	417
Db	1547	GCTTTTAAACCTCTCTCATCTCTTACTACATATTAAGAAATTCATCTCTGGGAGAAACCC	1608
QY	418	TTAAAGTGTAAACATGTATTAAGATTTCCTGCACTTCMAAATTAAGCAACAATTA	477
Db	1607	TACAAATGTGAAGATGTGGCAAAAGCTTTTAGCGGATCTCAAAACTTACGAACATTAAG	1666
QY	478	TTAAAACATCATGTGTG-GATCTCTGTGTTATCAATGTGATCATCTGTGTTTAAAAA	536
Db	1667	ATAATTTCACTACGTAGAGAAACCTTATTAAGTGAAGATGTGCAAAAGCTTTTAACCA	1728
QY	537	TTTCCAAACTGTGTACGATATTAACAATTCATATTAACAACCTGTGATCCAAACTTAAATG	596
Db	1727	TCTGCAAACT-----TACTTAACATTAATAAATACACTACTGGAGAGAACTACAGAC	1788
QY	597	TCTTAATGTGTAAAGGTGTGTGTGGAAAAAGTTTATCTTACATATGTTAAGTCA	656
Db	1781	TGGAATGTGTATATGATTTTGAACAACCTCAAAATTTTCTTAATATTAAGAAATCA	1848
QY	657	TGATGAT 663	
Db	1841	TACTGCT 1847	

RESULT 7
US-10-104-047-622
; Sequence 622, Application US/10104047
; Publication No. US20030236392A1

```
; GENERAL INFORMATION:
; APPLICANT: HELIX RESEARCH INSTITUTE
; TITLE OF INVENTION: No. US2003023692A1el full length cDNA
; FILE REFERENCE: H1-A0105
; CURRENT APPLICATION NUMBER: US/10/104,047
; PRIOR FILING DATE: 2002-03-25
; PRIORITY APPLICATION NUMBER:
; NUMBER OF SEQ ID NOS: 4096
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 622
; LENGTH: 3078
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-104-047-622

Query Match      7.4%; Score 91.8; DB 17; Length 3078;
Best Local Similarity 52.1%; Pred. No. 1.3e-07;
Matches 257; Conservative 0; Mismatches 227; Indels 9; Gaps 2;

QY      11 GTGACGAAACCAATGATGATATCATCTTAAATCTTCTCTTCTCATCGTCCCAAAA 70
DB      650 GTGGCAATATCATTTTGCATGCTTTCACATTAATCAACATAGAAAATTCATATCTAGAG 709
QY      71 AGTATATTTGCACATATGAAAGGTGATGATAAGCCATTAATCGACATATTATTAGAGC 130
DB      710 AGTATTTTCAATATGGAAGAAATGTGTAAAGCTTTAACTGTCTCTCAACCTTACTA 769
QY      131 AACATTTAAGAACCCAGATTAATGATGACCCGTAATAATGTACAGTGGACGATTGTGATA 190
DB      770 AACATTAATATTCATCTAGTGAAGAAAACCCCTCAAAATGT-----GAAGAAATGTGGCA 823
QY      191 AAGATTTTTCAGAAAATTCACATTTGGAAAACATATTGTAATCATCATTTCCGAAAAAAAC 250
DB      824 AAGCTTTTAAACCGGTCTCAAAATCTTACTAAACATTAATAATTCATATCGAGAGAAAC 883
QY      251 CATTCATATGTCAGTGTGTGTAAGGGGTTAATCTCGACAACACTTGGAAAAAGACATG 310
DB      884 CTRCAAAATGTGAAGATGTGGCAAGCTTTTAAACCGGTCTCTCAACCTTACTTAACATA 943
QY      311 AAATCACCATACAAAGTCAATTAATGTACATTTGAAAAATGTCAAGAAACATTTTATA 370
DB      944 AAAGAAATTCATACGAAGAGAAACCCCTAACAAATGTGAAGATGTGGCAAGCCCTTAAC 1003
QY      371 AACATCATCTTT---AAGACATATATATTTATGTCTTCATGAAAAAATTAACGTGTA 427
DB      1004 AGTTCGATCTTAAATTAACATTAAGAAATTCATATGAAAGATTAACCCCTACAAATGTG 1063
QY      428 AACATGTAATAAGTTTCTACCTGACCTTCAAAATTAAGCAACATTAATAATTAAATC 487
DB      1064 AAGAAATGTGGCAAGCCCTTTAGAGTATTTCTCAATTTCTTAATAAACATTAAGATTAATCA 1123
QY      488 ATGCTGATCTCC 500
DB      1124 CTGGGGAAGAAACC 1136

RESULT 8
US-10-094-749-692
; Sequence 692, Application US/10094749
; Publication No. US20030219741A1
; GENERAL INFORMATION:
; APPLICANT: ISOGAI, TAKAO
; APPLICANT: SUGIYAMA, TOMOYASU
; APPLICANT: OTSUKI, TETSUJI
; APPLICANT: WAKAMATSU, AI
; APPLICANT: SATO, HIROYUKI
; APPLICANT: ISHII, SHIZUKO
; APPLICANT: YAMAMOTO, JUN-ICHI
; APPLICANT: ISONO, YUUKO
; APPLICANT: HIO, YURI
; APPLICANT: OTSUKA, KAORU
; APPLICANT: NAGAI, KEIICHI
```

```
; APPLICANT: IRIE, RYOTARO
; APPLICANT: TAMECHIKA, ICHIRO
; APPLICANT: SEKI, NAOHICO
; APPLICANT: YOSHIKAWA, TSUTOMU
; APPLICANT: OTSUKA, MOTOTYUKI
; APPLICANT: NAGAHARI, KENJI
; APPLICANT: MASUHO, YASUHIKO
; TITLE OF INVENTION: NOVEL FULL-LENGTH CDNA
; FILE REFERENCE: 08435/0160
; CURRENT APPLICATION NUMBER: US/10/094,749
; PRIOR FILING DATE: 2002-03-12
; PRIOR APPLICATION NUMBER: 60/350,435
; PRIOR FILING DATE: 2002-01-24
; PRIOR APPLICATION NUMBER: JP 2001-328381
; PRIOR FILING DATE: 2001-09-14
; NUMBER OF SEQ ID NOS: 3381
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 692
; LENGTH: 2114
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-094-749-692

Query Match      7.3%; Score 90.4; DB 17; Length 2114;
Best Local Similarity 51.8%; Pred. No. 2.1e-07;
Matches 258; Conservative 0; Mismatches 231; Indels 9; Gaps 2;

QY      7 GAAAGTGACGAAACCAATGATGATATCATCTTAAATATCTTCTTCTTCTCATCGTCC 66
DB      1024 GAAATGTGGCAAGCCCTTCTCAATCTTCTCACCTTACTACATTAAGTAATTCATATCT 1083
QY      67 AAAAGTATTTTGCACATATGAAAGGTGATGATAAGCCATTAATCGACATCTATTATA 126
DB      1084 GGAAGAGAGCCCTTCAATATGTGAAGAAATGTGTAAAGCTTTAAACCCCTTACGCCCTT 1143
QY      127 GAGCAACATTTTAAAGAACCCAGATATGATGACCGTATTAATGTATGATGACATGAGATGT 186
DB      1144 ACTACACATAGTTCATCTATCTATGTTAAAGAAAACCCCTCAAAATGT-----GAAGAAATGT 1197
QY      187 GATTAAGCATTTTTCAGAAAATTCACATTTGAAAAACATATTTGATATCATTTCCGAAAAA 246
DB      1198 GACAAAGCTTTTAAACCATTTCTATCTTACTTAACATTAAGTAATTCATATCTGAGAG 1257
QY      247 AAACATTCATTTGATGATGTGTGTAAGGGGTTAATCTCGACAACCTTGAAGAAGA 306
DB      1258 AAATCTTACAAATGTGAACATGTGGCAAGGCTTTAAGCTGTCTTCAACCTTACAAA 1317
QY      307 CATGAATTCACCAATACAAAGTCAATTAATGTATCATTTGAAAAATGTCAAGAAAGCATTT 366
DB      1318 CATGAAGAAATTCATCTATCTGAGAGAAACCCCTACAAATGTGAAGATGTGGCAAGCCCTT 1377
QY      367 TATTAACAT--CAATCTTAAAGCATATATATATCTGTTCAAGAAAAAATTAAGC 423
DB      1378 AATGTGCTTACACCTTACTACATTAAGATGATTCATCTGAGAGAAACCCCTACAAA 1437
QY      424 TGTAAACAATGTATAAGTTTTCATCTGACCTTCAAAATTAAGCAACAATAATTATAA 483
DB      1438 TGTGAAGAAATGTGGCAAGGCTTTTAAACCACTCTCAAAACCTTACTATATCAATTAATTA 1497
QY      484 CATCATGTGATCTCTCT 501
DB      1498 CATCTGGAAGAAACCT 1515

RESULT 9
US-10-094-749-499
; Sequence 499, Application US/10094749
; Publication No. US20030219741A1
; GENERAL INFORMATION:
; APPLICANT: ISOGAI, TAKAO
; APPLICANT: SUGIYAMA, TOMOYASU
; APPLICANT: OTSUKI, TETSUJI
; APPLICANT: WAKAMATSU, AI
```

```
/ APPLICANT: SATO, HIROYUKI
/ APPLICANT: ISHII, SHIZUKO
/ APPLICANT: YAMAMOTO, JUN-ICHI
/ APPLICANT: ISONO, YUUKO
/ APPLICANT: HIO, YURI
/ APPLICANT: OTSUKA, KAORU
/ APPLICANT: NAGAI, KEIICHI
/ APPLICANT: IRIE, RYOTARO
/ APPLICANT: TAMECHIKI, ICHIRO
/ APPLICANT: SEKI, NAOHICO
/ APPLICANT: YOSHIKAWA, TSUTOMU
/ APPLICANT: OTSUKA, MOTOMYUKI
/ APPLICANT: NAGAHARI, KENJI
/ APPLICANT: MASUHO, YASUHIKO
/ TITLE OF INVENTION: NOVEL FULL-LENGTH cDNA
/ FILE REFERENCE: 084335/0160
/ CURRENT APPLICATION NUMBER: US/10/094,749
/ PRIOR FILING DATE: 2002-03-12
/ PRIOR APPLICATION NUMBER: 607/350,435
/ PRIOR FILING DATE: 2002-01-24
/ PRIOR APPLICATION NUMBER: JP 2001-328381
/ PRIOR FILING DATE: 2001-09-14
/ NUMBER OF SEQ ID NOS: 3381
/ SOFTWARE: PatentIn Ver. 2.1
/ SEQ ID NO 499
/ LENGTH: 2230
/ TYPE: DNA
/ ORGANISM: Homo sapiens
US-10-094-749-499
```

```
Query Match 7.3%; Score 90.2; DB 17; Length 2230;
Best Local Similarity 50.4%; Pred. No. 2.4e-07;
Matches 332; Conservative 0; Mismatches 313; Indels 14; Gaps 4;
```

```
QY 7 GAAAGTGAAGAAACCAATCGATATCATCTTTAATCTTTCTTCTTCATACGCTCC 66
DB 656 GAATGTGAACAAATCACTTGGCATGCTTTTCAAGCTTCAACATATAAAATTCATACT 715
QY 67 AAAAGTATATTTGACATATGAAAGGTGTGATTAAGCCTATATGACCATCATATTA 126
DB 716 AGAGAAATTTCTACAAATGTGAAGGTGTGAAAAACCTTTAAGTGTGTCAAACTT 775
QY 127 GAGCAACATTTAAGAAACCAAGTATGATCGACCGTATTAATGTACAGTGAAGATGT 186
DB 776 TCTAACTTAAGAAATTCATACCTGAGAAAAACCTTACAAATGTGAAGTA-----TGT 829
QY 187 GATTAAGCATTTTTCAGAAAAATCACATTGGAACCATTTGTATCATCTCCGAAAAA 246
DB 830 GGAAGAGCCTTTTCAACAAATCTTAACTTAACTTAAGATTAATGTGTACTGAGAA 889
QY 247 AAACATTCATTTGTTCAAGTGTGTGTAAGGGGTAAATTTCTGCAACAACCTTGAAAGA 306
DB 890 AAACCTTAATAATGTGACACCTGTGCAAAACCTTTTAACAGTCCCTTACCTTACTGA 949
QY 307 CATGAATCAACCCATACAAAGTCATTTAAATGTACATTTGAAAAATTTGTCAAGAGCATTT 366
DB 950 CATTAATATATCTACTGTAAGAGAAACCTTCAAAATGTGAACAATGTGGCAGTCTTT 1009
QY 367 TATTAACATCAATC---TTTAAGACATCATATATATCTGTTCATGAAAAAACATTAAG 423
DB 1010 AAGCAGTCCCAACCCCTTCAACATCAAGATTAATTAATCTGGAAGGAAACCATTCMAA 1069
QY 424 TGTAAACAATGTATTAAGTTTCACTCGACCTTCAAAATTTAGCACAACAATAATTAATAA 483
DB 1070 TGTGAGAAATGTGGCAAGCTTTTAACCTATCTTAACTTAATCTGAACTTAAGAAAAAT 1129
QY 484 CATCATGATG-GATCTCTGCTCTTATCAATGTGATCATCTGTGTGTTTAAATAATTTCA 542
DB 1130 TACACTAGAGAGAAAGCTTAACAATGTGAAGATGTGGCAAGCCTTTAAACAGTTTGA 1189
QY 543 AACTGGTCAATTAATTAATTTTCAATTTTCAATTAACAACTGCATCCAAAATTATGTCTTA 602
DB 1190 ACCCTATTATACATTAAGATTAATTCATAGCAGAGAGAAACCC---CACAAATGTGAAGA 1245
```

```
QY 603 ATGTGTAAAGGTGTGTGGAAAAAGTTTATCTTCATATATGTTAAGTCATGATG 661
DB 1246 ATGTGGACAGCTTTTAACCAAGTCCGCAAAAGCTCACTGAACATTAATTCATACTG 1304
```

RESULT 10

```
US-10-029-386-13298
/ Sequence 13298, Application US/10029386
/ Publication No. US20030194704A1
/ GENERAL INFORMATION:
/ APPLICANT: Penn, Sharon G.
/ APPLICANT: Rank, David R.
/ APPLICANT: Hanzel, David K.
/ TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
/ FILE REFERENCE: AEWICA-X-2
/ CURRENT APPLICATION NUMBER: US/10/029,386
/ PRIOR FILING DATE: 2001-12-20
/ NUMBER OF SEQ ID NOS: 34288
/ SOFTWARE: Annomax Sequence Listing Engine ver. 1.1
/ SEQ ID NO 13298
/ LENGTH: 590
/ TYPE: DNA
/ ORGANISM: Homo sapiens
/ FEATURE:
/ OTHER INFORMATION: MAP TO CHR19.3
/ OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 4
/ OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 4.5
/ OTHER INFORMATION: NT HIT: g11475315, EVALUE 0.00e+00
/ OTHER INFORMATION: SWISSPROT HIT: Q05481, EVALUE 1.00e-119
/ OTHER INFORMATION: EST_HUMAN HIT: BG619358.1, EVALUE 0.00e+00
US-10-029-386-13298
```

```
Query Match 7.3%; Score 89.8; DB 16; Length 590;
Best Local Similarity 53.8%; Pred. No. 1.7e-07;
Matches 224; Conservative 0; Mismatches 192; Indels 9; Gaps 2;
```

```
QY 62 GTCCCAAAAAGTATTTTGCATATGAAAGGTGTGATTAAGCCTTAATGACCATCAT 121
DB 126 GTGCTTAAGAGAAATCTACAGTGTGAAGATGTGGCAAGCATTTCTATGTCTCTCA 185
QY 122 TATTAGAGCAACATTTAAGAACCCACAGTATGATGACCGTATTAATGTACAGTGA 181
DB 186 CCTTAATAGACATTAAGAGATACACACTGAGAGAAACCTTAACAATGT-----GAAG 239
QY 182 ATTTGATTAACATTTTTCAGAAAAATCACATTGGAACCATATTTGATCATCTCCG 241
DB 240 AATGTGGCAACCTTTTGAAGCCATCTTCAACCTTGTCTAAACATTAAGAGATTCATCTG 299
QY 242 AAACAAACCATTCATTTGTTCAAGTGTGTGTAAGGGGTAAATTTCTGACAAACCTTGA 301
DB 300 GAGAGAAACCTTCAACAATGTGAAGATGTGGCAAGCTTTTGAAGCCGTTTCAACCTTG 359
QY 302 AAAGCATGAATCAACCCATACAAAGTCATTTAAATGTACATTTGAAAAATTTGTCAAG 361
DB 360 CTAAACATTAAGAAATTCATCTGAGAGAAACCTTCAAAATGTAAAGATGTGGCAAG 419
QY 362 CATTTTATTAACATCAATTTTAAAGACATCATATATATCTGTTCATGAA---AAACAT 418
DB 420 CTTTAAAGCAATTCCTCAACCTTGTCTTAATCAAGTAACTATCTGTAAGAGAAACCTT 479
QY 419 TAACGTGTAAACAATGTATTAAGTTTCACTCGACCTTCAAAATTTAGCACAACAATAAT 478
DB 480 ACAAATGTAAAGATGTGCAAGCTTTTAAAGCACTCTCAACCTTACTTAACAATTAATA 539
QY 479 TAAAAATCATGATG 493
DB 540 TAATCATGCTGAG 554
```

```
RESULT 11
US-10-108-260A-29
```

```
; Sequence 29, Application US/10108260A
; Publication No. US20040005560A1
; GENERAL INFORMATION:
; APPLICANT: HELIX RESEARCH INSTITUTE
; TITLE OF INVENTION: No. US20040005560A1 full length cDNA
; FILE REFERENCE: H1-A0106
; CURRENT APPLICATION NUMBER: US/10/108,260A
; CURRENT FILING DATE: 2002-03-27
; NUMBER OF SEQ ID NOS: 5458
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 29
; LENGTH: 2221
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-108-260A-29

Query Match      7.2%; Score 89.2; DB 17; Length 2221;
Best Local Similarity 53.4%; Pred. No. 3,7e-07;
Matches 237; Conservative 0; Mismatches 198; Indels 9; Gaps 2;

QY 61 CGTCCCAAAAGTATATTGACATATGAAAGGGTGTATAAAGCTTATTCAGACATCA 120
DB 654 CATACCGAGAGAAACCTTTCAGATGTAGAAATGGCAAAAGCTTTAATAGAGCTCA 713
QY 121 TTAATTAGAGCAACATTTAAGAACCCACAGTAATGATGACCGTATAAATGACGTGAC 180
DB 714 TAGCTTACTAGCATCAGAGATGATCTGAGAGAAACCTTTCAAGTGTGTA-----AG 767
QY 181 GATTGTGATTAAGCAATTTTTCAGAAATCAATTTGGAAACATATTTGATCATTCC 240
DB 768 GAATGTGGCAAAAGCTTTTAAACAGAGCTTCAACCTTACTCAACATCAGAAATCCATCT 827
QY 241 GAAAAAAACCATTCCATTGTTCAGTGTGTGTAAAGGGTAAATTTTCAGCAACACTTG 300
DB 828 GAGAGAGAAACCTTCAATGTAAAGATGTGGCAAAACCTTTTAAACAGGGCTCAACCTT 887
QY 301 AAAAGACATGAATCAACCCATACAAAGTCAATTTAATGTACATTTGAAAAATGTCAAGAA 360
DB 888 ACTCGACATCAAGAAATCCATCTGAGAGAGAAAGCTTTCAAAATGTGAAGATGTGGCAAA 947
QY 361 GCATTT---TATTAACATCAATCTTTAAGACATCATATATTTCTGTTCAAGAAAAACA 417
DB 948 GCTTTTATCTGGGGCTCAACCTTACTCAACATCAGAGAGATCCATCTAGGAGAAATTC 1007
QY 418 TTAAAGGTAAACATGTAATTAAGTTTTCATCTGACCTTCAAAATTAAGCAACATATAA 477
DB 1008 TTCAAATGTAAAGATGTGGCAAAAGCTTTTAAACGAGAGCTCACACCTTACTCAACATCAG 1067
QY 478 TTAAACATCATGTTGATCTCCT 501
DB 1068 AGAATTCATCTGAGAGAAACCT 1091

RESULT 12
US-10-172-118-910
; Sequence 910, Application US/10172118
; Publication No. US2003022437A1
; GENERAL INFORMATION:
; APPLICANT: Dai, Hongyue
; APPLICANT: He, Yudong
; APPLICANT: Linsley, Peter
; APPLICANT: Mao, Mao
; APPLICANT: Roberts, Chris
; APPLICANT: Van 't Veer, Laura
; APPLICANT: Van de Vijver, Marc
; APPLICANT: Bernades, Rene
; TITLE OF INVENTION: Diagnosis and Prognosis of Breast Cancer Patients
; FILE REFERENCE: 9301-175-999
; CURRENT APPLICATION NUMBER: US/10/172,118
; CURRENT FILING DATE: 2002-06-14
; PRIOR APPLICATION NUMBER: 60/380,770
; PRIOR FILING DATE: 2002-05-14
; NUMBER OF SEQ ID NOS: 2699
```

```
; SEQ ID NO 910
; LENGTH: 3839
; TYPE: DNA
; ORGANISM: Homo sapiens
; PUBLICATION INFORMATION:
; DATABASE ACCESSION NUMBER: NM_003430
; DATABASE ENTRY DATE: 2001-06-18
US-10-172-118-910

Query Match      7.2%; Score 89.2; DB 17; Length 3839;
Best Local Similarity 53.7%; Pred. No. 4.6e-07;
Matches 234; Conservative 0; Mismatches 193; Indels 9; Gaps 2;

QY 69 AAAGTATTTTGCATATGAAAGGGTGTATAAAGCCTTAATTCAGACCATTAATTATAG 128
DB 964 AGAGAAACCTTACAAATGTGAAGATGTGGCAAAAGCTTTTAGCATTTTCAACCTTGC 1023
QY 129 GCAACATTTAAGAACCCACAGTAATGATGACCGTAAATGTAACAGTGCAGATTGTGA 188
DB 1024 TAAACATTAAGAAATTCATATCTGAGAGAAACCTTCAAAATGT-----GAAGATGTGG 1077
QY 189 TAAAGCATTTTTCAGAAATTCATTTGGAAACATATTTGATCTCATTTCCGAAAAAAA 248
DB 1078 CAAAGCTTTTACGCAATCTTTCAGCCCTTGTAAACATTAAGAAATTCATATCTGAGAGAA 1137
QY 249 ACCATTCCATTTGTTCAGTGTGTGTAAAGGGTTAATTTCCGAAACACTTGAAGAGACA 308
DB 1138 ACCCTTCAAAATGTAAAGATGTGGCAAAAGCTTTTACAAATTTCTCAACCTTCTAATCA 1197
QY 309 TGAATACCCATACAAAGTCAATTTAATGTACATTTGAAATTTGTCAAGAAACATTTTA 368
DB 1198 TAGATTAATCTACATCAGAGAGAAACCTTCAAAATGTAAAGATGTGACAAATCTTTTA 1257
QY 369 TAAACATCAATC---TTTAAAGACATATATTTATCTGTTGATGAAAAAACAATTACGTG 425
DB 1258 GCGACTCTCAACCTTACTTAACATTAATTAATCAATGTGCGAGAGAAACCTTCACAANTG 1317
QY 426 TAAACATGTAATTAAGTTTTCATCTGACCTTCAAAATTAAGCAACATTAATTAACA 485
DB 1318 TGAAGATGTGGCAAAAGCTTTTAAATGATCTTCAAAATCTTACTATACATTAAGTTTATCA 1377
QY 486 TCATGTTGATCTCCT 501
DB 1378 TACTGAGAGAAACCT 1393

RESULT 13
US-10-342-887-910
; Sequence 910, Application US/10342887
; Publication No. US20040058340A1
; GENERAL INFORMATION:
; APPLICANT: Dai, Hongyue
; APPLICANT: He, Yudong
; APPLICANT: Linsley, Peter S.
; APPLICANT: Mao, Mao
; APPLICANT: Roberts, Christopher J.
; APPLICANT: Van 't Veer, Laura Johanna
; APPLICANT: Van de Vijver, Marc J.
; APPLICANT: Bernades, Rene
; TITLE OF INVENTION: Diagnosis and Prognosis of Breast Cancer Patients
; FILE REFERENCE: 9301-188-999
; CURRENT APPLICATION NUMBER: US/10/342,887
; CURRENT FILING DATE: 2003-01-15
; PRIOR APPLICATION NUMBER: 60/298,918
; PRIOR FILING DATE: 2001-06-18
; PRIOR APPLICATION NUMBER: 60/380,710
; PRIOR FILING DATE: 2002-05-14
; PRIOR APPLICATION NUMBER: 10/172,118
; PRIOR FILING DATE: 2002-06-14
; NUMBER OF SEQ ID NOS: 2699
; SEQ ID NO 910
; LENGTH: 3839
; TYPE: DNA
```

```
; ORGANISM: Homo sapiens
US-10-342-887-910
Query Match      7.2%; Score 89.2; DB 17; Length 3839;
Best Local Similarity 53.7%; Pred. No. 4.6e-07;
Matches 234; Conservative 0; Mismatches 193; Indels 9; Gaps 2;

QY      69  AAGATATATTTGACATATGAGAGGGTGTGATTAAGCCCTATATATGACCATATTATTAGA 128
DB      964  AGAGAAACCTTACAAATGTGAAGAAATGTGGCAAAAGCTTTAGCCATTCTTCAACCTTGC 1023

QY      129  GCAACATTTAAGAACCCACAGTAATGACCCGTATTAATGACAGTGGACGATTGTGA 188
DB      1024  TAAACCTAAGAGATTCATCTGAGAGAAACCTTCAAAATGT-----GAAGAAATGTG 1077

QY      189  TAAAGCATTTTTCAGAAAATTCACATTTGGAACACATTTGTATTCATTTCCGAAAAAAA 248
DB      1078  CAAAGCTTTTTCAGCATTTCTTCAAGCCCTTGTCTAAACATTAAGAAATTCATCTGAGAGAA 1137

QY      249  ACCATTCATTTGTCAGTGTGTGTAAAGGGTTAATTCGACAAACCTTGAAGAAAGACA 308
DB      1138  ACCCTCAAAATGTAAAGAAATGTGGCAAAAGCTTTTACAAATTCCTCAACCTTGTATCA 1197

QY      309  TGAATACCCCATACAAAGTCAATTTAAATGTACATTTGAAAATTTGCAAGAAAGATTTA 368
DB      1198  TAAGATTAATCTCATCTGAGAGAAAGAAACCTTACAAATGTAAAGATGTGACAAACCTTTAA 1257

QY      369  TAAACATCATC---TTTAAAGACATCATATATATCTGTTCAATGAAAAACATTAACTG 425
DB      1258  GGCATCTCAACCTTCACTAATCAATTAATTAATCACTGAGAGAAACCTTCAAAATG 1317

QY      426  TAAACATGTATTAAGATTTTCACTGACCTTCAAAATTTAGCAACATTAATTAAGACA 485
DB      1318  TGAAGATGTGGCAAGCTTTTATCGATCTTCAATCTTACTATACATTAAGTTATTGA 1377

QY      486  TCATGTGTGATCTCTCT 501
DB      1378  TACTGTGAGAAACCT 1393

RESULT 14
US-10-098-841-48
; Sequence 48, Application US/10098841
; Publication No. US20020197679A1
; GENERAL INFORMATION:
; APPLICANT: Tang, Y. Tom
; APPLICANT: Liu, Chenghua
; APPLICANT: Asundi, Vinod
; APPLICANT: Xu, Chongjun
; APPLICANT: Zhou, Ping
; APPLICANT: Ma, Yungqing
; APPLICANT: Wang, Jian-Rui
; APPLICANT: Zhao, Qing A.
; APPLICANT: Ren, Feiyan
; APPLICANT: Chen, Rui-hong
; APPLICANT: Wang, Dunrui
; APPLICANT: Wang, Zhiwei
; APPLICANT: Wehrman, Tom
; APPLICANT: Zhang, Jie
; APPLICANT: Qian, Xiaohong B.
; APPLICANT: Drmanac, Radoje T.
; TITLE OF INVENTION: No. US20020197679A1el Nucleic Acids and
; FILE REFERENCE: Polypeptides
; FILE REFERENCE: 784CIP2
; CURRENT APPLICATION NUMBER: US/10/098,841
; CURRENT FILING DATE: 2002-03-13
; PRIOR APPLICATION NUMBER: 09/598,042
; PRIOR FILING DATE: 2000-06-20
; PRIOR APPLICATION NUMBER: 09/552,317
; PRIOR FILING DATE: 2000-04-25
; PRIOR APPLICATION NUMBER: 09/488,725
; PRIOR FILING DATE: 2000-01-21
; NUMBER OF SEQ ID NOS: 331
```

```
; SOFTWARE: pf_fl_genes Version 1.0
; SEQ ID NO 48
; LENGTH: 3502
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (154)..(2583)
; NAME/KEY: misc_feature
; LOCATION: (1)..(3502)
; OTHER INFORMATION: n = a, t, c or g
US-10-098-841-48
Query Match      7.2%; Score 89; DB 13; Length 3502;
Best Local Similarity 51.7%; Pred. No. 4.8e-07;
Matches 256; Conservative 0; Mismatches 230; Indels 9; Gaps 2;

QY      79  TGCATATGAAGGGTGTGATTAAGCCATATATGACCATCTATTATGAGCAACATTTA 138
DB      2101  TACAAATGTGAAGAAATGTGGCAAAAGCTTTAAGTGTCTCAACCTTAACTAAACATAG 2160

QY      139  AGAACCCACAGTAATGATCGACCGTATTAATGTACAGTGAAGATTTGATTAAGCATTT 198
DB      2161  ATTAATTCATCTGAGAGAAACCTTACAAATGT-----GAAGATGTGGCAAGCTTTT 2214

QY      199  TTCAGAAAATCACATTTGGAACACATATTTGATCATTCGAAAAAAAACCATTCAT 258
DB      2215  AAACGTCTCAACCTTTCTACATATAGATTAATTCATCTGAGAGAAACCTTACAAA 2274

QY      259  TGTTCAGTGTGTGTAAAGGGGTTAATCTTGACAAACCTTGAAGAAAGCATGAATCACC 318
DB      2275  TGTGAAAAATGTGGCAAGCTTTTAAACGACCTTCAACCTTATTTGAACATTAAGAAAT 2334

QY      319  CATCAAAAGTCATTAATTAATGTACATTTGAAAATTTGCAAGAAAGATT---TATTAACAT 375
DB      2335  CATCTGAGAGAACCTTACAAATGTGAAGATGTGGCAAGCATTTTACATATTCCTCA 2394

QY      376  CATCTTAAGACATCATATATATCTGTTCATGAAAAAACATTAAAGTGAACATATGT 435
DB      2395  CACCTTAATACATTAAGAAATTCATCTAAAGAGCAACCTTACAAATGTAAAGATGT 2454

QY      436  AATAAAGTTTTCACCTCGACCTTCAAAATTTAGCAACAATTAATTAACATATGTGCA 495
DB      2455  GGCAAAGCTTCAACCAATATTTCAACCTTACACATTAACAAATTCATCTGAGAGAG 2514

QY      496  TCTCCTGTATCATATGATGATCATCTGTTGTTTAAAAATTTCCAAACTGTGCAGTA 555
DB      2515  AAACCTTACAAACCTGAAAGATGTGACATGATTTTGACACACCTTCAACCTTTTCAAC 2574

QY      556  TTACATTTTCATATA 570
DB      2575  ATTAATAATTAATATA 2589

RESULT 15
US-10-029-386-25074
; Sequence 25074, Application US/10029386
; Publication No. US20030194704A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharron G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR G
; FILE REFERENCE: ABOVICA-X-2
; CURRENT APPLICATION NUMBER: US/10/029,386
; CURRENT FILING DATE: 2001-12-20
; NUMBER OF SEQ ID NOS: 34288
; SOFTWARE: Anomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 25074
; LENGTH: 1466
; TYPE: DNA
; ORGANISM: Homo sapiens
```

```
; FEATURE:
; OTHER INFORMATION: MAP TO AC011467.3
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 0.67
; OTHER INFORMATION: NT HIT: g115309152, EVALU0.00e+00
; OTHER INFORMATION: SWISSPROT HIT: P35789, EVALU0.00e+00
; OTHER INFORMATION: EST_HUMAN HIT: AUI43734.1, EVALU0.00e+00
US-10-029-386-25074
```

```
Query Match 7.2%; Score 88.6; DB 16; Length 1466;
Best Local Similarity 53.3%; Pred. No. 4.1e-07;
Matches 237; Conservative 0; Mismatches 199; Indels 9; Gaps 2;
```

```
QY 60 ACCTCCCAAAAGTATATTTGCAATATGAAAGGATGATTAAGCTATATGACCATC 119
Db 272 ACATTAAGTGAAGAAATCTTCAATATGTAAGATGTAAGAAAGTCAATTTGCAATGCTTC 331
QY 120 ATTATTAAGCAACATTTAAGAACCAACAGTAATGATCGACGTATTAATGTACAGTGA 179
Db 332 ACACCTTAGCTCAACATTAAGAAATTCATAGTGAAGAAACCTACAAATGTAA----- 386
QY 180 CGATGTGTAAGCAATTTTTCAGAAATGCAATTTGAAACATATTTGTATCAGATTC 239
Db 387 -GATGTGGAAAGCCTATATATGAGACCTCAACCTTTCTACATTAAGAAATTCATAC 445
QY 240 CGAAAAAAACCATTCATTTGTCAGTGTGTGTAAGGGTTAATTTCTGACACACCTT 299
Db 446 TGGAAAGAAACCTTACAAATGCGAAGGTGTGAAAAAGCCTTTAACCGGCTCTCACACCT 505
QY 300 GAAAGACATGAATACCCCATACAAAGTCATTAAATGTACATTTGAAAAATTGTCAAGA 359
Db 506 TACTACACATTAAGTATATTCATCTGGAAGAAACCTCAAAATGTGAGAGGTGGCAA 565
QY 360 AGCATTTTAAACAT---CAATCTTAAGACATCATATTTATCTGTTGATGAAAAAAC 416
Db 566 AGCTTTTAACCAATCTGCAAAACCTTACTACATTAAGAGAAATTCATACCTGAGAAACC 625
QY 417 ATTACGTGTAAACATATGTAATTAAGTTTCACTGACCTTCAAAATTAGCACACATTA 476
Db 626 CTACAAATGTGAAGAAATGTGCGAGAGCTTTAGCCAGTCTCAACCTTTACTGCACTTA 685
QY 477 ATTAAACATCATGTGTGATCTCTCT 501
Db 686 GATTAATTCATGTGAGAGAAACCT 710
```

Search completed: February 9, 2005, 11:50:30
Job time : 2157.88 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 9, 2005, 01:45:39 / Search time 3806.38 Seconds
(without alignments)
12360.150 Million cell updates/sec

Title: US-09-831-804-1_COPY_720_1955
Perfect score: 1236
Sequence: 1 atgagtcgaagtcgaagcgaac.....aaacatcagttcttcga 1236

Scoring table: IDENTITY_NUC
Gapop 10.0, Gapext 1.0

Searched: 34239544 seqs, 19032134700 residues

Total number of hits satisfying chosen parameters: 68479088

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database :

EST:*
1: gb_est1:*
2: gb_est2:*
3: gb_est3:*
4: gb_est4:*
5: gb_est5:*
6: gb_est6:*
7: gb_est7:*
8: gb_est8:*
9: gb_est9:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	182.2	14.7	628	BZ298000	BZ298000 CG3957.f1
C 2	167.6	13.6	959	CNS060MP	AL405671 T7 end OE
3	103.8	8.4	801	CN759437	CN759437 ID0AAA25C
4	103.8	8.4	898	CN761003	CN761003 ID0AAA2AB
5	98.2	7.9	1609	AK032220	Mus muscu
6	97.8	7.9	2647	BC047646	BC047646 Homo sapi
7	97.6	7.9	745	AU123448	AU123448 Homo sapi
8	93.4	7.6	727	CO434402	CO434402 UT-W-HXO-
9	92.4	7.5	631	CN352904	CN352904 Homo sapi
10	92.4	7.5	2672	BC036394	BC036394 Homo sapi
11	92	7.4	591	AQ005136	AQ005136 CTF-HSP-2
12	92	7.4	687	CN362485	CN362485 170004241
13	92	7.4	752	CN417582	CN417582 170004240
14	91	7.4	593	B9387	B9387 CTF-HSP-228
15	91	7.4	820	BX414627	BX414627 Homo sapi
16	90.8	7.3	666	CN297781	CN297781 170005321
17	90.4	7.3	2694	BC022527	BC022527 Homo sapi
18	90.4	7.3	2698	BC037782	BC037782 Homo sapi
19	90.2	7.3	2253	BC028252	Mus muscu
20	89.8	7.3	350	BP292533	BP292533 Mus muscu
21	89.2	7.2	582	BP234906	BP234906 Homo sapi
22	89.2	7.2	604	BE161630	BE161630 MR3-HT044
23	89.2	7.2	638	CN341035	CN341035 170006000
24	89.2	7.2	673	CN353930	CN353930 170006001

25	89.2	7.2	3609	3	BC032590	BC032590 Homo sapi
26	89	7.2	2856	3	HSMB03730	AL832422 Homo sapi
27	88.6	7.2	535	1	AL705393	AL705393 DKFZ686M
28	88.6	7.2	612	5	BP238960	BP238960 BP238960
29	88.6	7.2	2174	3	AK033001	AK033001 Mus muscu
30	88.6	7.2	2330	3	BC037426	BC037426 Homo sapi
31	88.6	7.2	2622	3	AK033958	AK033958 Mus muscu
32	88.4	7.2	998	5	BX456765	BX456765 Mus muscu
33	88	7.1	895	5	BU195416	BU195416 AGENECOURT
34	87.8	7.1	608	7	CN266758	CN266758 170004554
35	87.8	7.1	4251	3	CR749856	CR749856 Homo sapi
36	87.6	7.1	744	5	BX437291	BX437291 BX437291
37	87.2	7.1	579	8	AQ347265	AQ347265 RPI11-11
38	87.2	7.1	687	8	AQ389266	AQ389266 RPI11-15
39	87.2	7.1	907	5	BO423752	BO423752 AGENECOURT
40	87	7.0	660	6	CD770140	CD770140 AGENECOURT
41	87	7.0	842	5	BU507408	BU507408 AGENECOURT
42	86.8	7.0	521	8	AQ475626	AQ475626 CITBI-E1-
43	86.8	7.0	571	7	CN280110	CN280110 170006001
44	86.8	7.0	705	8	AQ194282	AQ194282 RPI11-61
45	86.8	7.0	860	8	AQ749175	AQ749175 HS_5575_A

ALIGNMENTS

RESULT 1
BZ298000/c 628 bp DNA linear GSS 31-OCT-2002
LOCUS CG3957.f1 Candida glabrata Random Genomic Library Candida glabrata
DEFINITION genomic clone CG3957, genomic survey sequence.
ACCESSION BZ298000
VERSION BZ298000.1 GI:24440936
KEYWORDS GSS.
SOURCE Candida glabrata
ORGANISM Candida glabrata
Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
Saccharomycetales; mitosporic Saccharomycetales; Candida.
REFERENCE 1 (bases 1 to 628)
AUTHORS Wong,S., Fares,M.A., Zimmermann,W., Butler,G. and Wolfe,K.H.
TITLE Evidence from comparative genomics for a complete sexual cycle in the 'asexual' pathogenic yeast Candida glabrata
JOURNAL MEDLINE 22509158
PUBMED 12620120
COMMENT Contact: Wong S
Department of Genetics, Smurfit Institute
Trinity College Dublin
Dublin 2, Ireland
Tel: 353 1 6082319
Fax: 353 1 6798558
Email: swong@tcd.ie
Class: plasmid ends.
Location/Qualifiers
FEATURES
source 1..628
/organism="Candida glabrata"
/mol_type="genomic DNA"
/strain="CBS 138"
/db_xref="taxon:5478"
/clone="CG3957"
/clone_1b="Candida glabrata Random Genomic Library"
ORIGIN
Query Match 14.7%; Score 182.2; DB 8; Length 628;
Best Local Similarity 59.5%; Pred. No. 7.3e-26;
Matches 348; Conservative 0; Mismatches 228; Indels 9; Gaps 2;
QY 84 ATTTGAAGGCTGATTAAGCCCTTAATGACCATCATTTTGAAGCAATTAAGAAC 143
DB 625 ATTCGATTAATTGGACAAAGCCCTTTCAAGGCCGCTTGTCTACAGAACACAGAAATAC 566
QY 144 CCACAGTAATGATCGACCGCTAATAATGTACAGTGACGAGATTGTGATTAAGCATTTTTCAG 203

Db 565 ---AGTCATCTGGAGGAGAAACCTTGAAAGTAAATCAATGTGAAGTTCATTACTAA 509

Qy 204 AAAATCACATTTGGAAACACATATTTGTATCATCTCCGAAAAAACCATTCCATTGTTTC 263

Db 508 AAAAGATCCACTTAGAGAGACCTTATACACACACAGATGAAAGACCGTTTATTTGTTTC 449

Qy 264 AGTGTGTGTAAAGGGGTTAATTCGACACACCTTGAAAGACATGAAATCAACCCATAC 323

Db 448 ATTTGTGGGAAAGGGGCTAATTAATCTAGGCAACAACTGAAACGATAGAGTAATCTACAC 389

Qy 324 AAAGCATTTAATGTACATTTGAAATTTGTCAAGAGATTTTAAATCAATCAATCTTT 383

Db 388 CAATGCTTCAATGTGAATATGAAATGATGAGTTCTACAGACACCTTCAAT 329

Qy 384 AAGACATCATATATATCTGTTTCATGAAAAAATTATGAGTGAACAATATTAAGT 443

Db 328 AAGGCTCATATTTTAGAGTTTCATTTACAAAGTCTAAATGTCAGATGACAAAG 269

Qy 444 TTTTCACTGACCTTCAAAATTAAGCACATATTAATTAATCAAT-----GGTGGATC 497

Db 268 TTTTCAAAACCTTACAGACTCAAAAATGACATAGCAAAACATATATCCAGATTTGT 209

Qy 498 TCCTGCTTATCATGTGATCATCTGTTGTTTAAATTTCCAACTGTCAGTAT 557

Db 208 CATGCTTATCATGTATCTTCACTTCACTGTTGTTCTAAGAGTTCAAAACATGCTGCTT 149

Qy 558 ACAATTTCAATATAAACAATGATCCAACTCAAACTTAATGCTTAATGTGTAAGTTG 617

Db 148 AAGATTACATGTTAAATAATGATCATCCGAAGTTAAATCCCTATATGACGTAACCTCG 89

Qy 618 TGTGGGAAAAAGGTTTATCTTCACATATGTTAAGTCATGATGA 662

Db 88 TGTGGGGAAGCGGTTTAAATATGACATGAAATCCACAGATGA 44

RESULT 2
CNS06MMP/c
LOCUS
DEFINITION
T7 end of clone AU0A005F10 of library AU0AA from strain CBS 3082
of *Saccharomyces kluyveri*, genomic survey sequence.
ACCESSION
AL405671
VERSION
AL405671.1 GI:12168715
KEYWORDS
GSS.
SOURCE
Saccharomyces kluyveri
ORGANISM
Saccharomyces kluyveri
Bukhryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
Saccharomycetales; Saccharomycetaceae; Saccharomyces.

REFERENCE
AUTHORS
1 (bases 1 to 959)
Soucie,J.L., Aigle,M., Artiguenave,F., Blandin,G.,
Bolotin-Fukuhara,M., Bon,E., Brothier,P., Casaregola,S.,
de-Montigny,J., Dujon,B., Durrens,P., Lepingle,A., Llorente,B.,
Malbertuy,A., Neugeglise,C., Olier-Kalogeropoulos,O., Potier,S.,
Sautin,W., Tekra,F., Toffano-Nicche,C., Wesolowski-Louvel,M.,
Winkler,P. and Weissenbach,J.
Genomic exploration of the hemiascomycetous yeasts: 1. A set of
yeast species for molecular evolution studies
FEBS Lett. 487 (1), 3-12 (2000)

REFERENCE
AUTHORS
TITLE
2 (bases 1 to 959)
Neugeglise,C., Bon,E., Lepingle,A., Winkler,P., Artiguenave,F.,
Galliardin,C. and Casaregola,S.
Genomic exploration of the hemiascomycetous yeasts: 9.
Saccharomyces kluyveri
FEBS Lett. 487 (1), 56-60 (2000)

JOURNAL
MEDLINE
20584719
11152876
11152884
3 (bases 1 to 959)
Genoscope.
Submitted (07-SEP-2000) Genoscope - Centre National de Sequencage,
2 rue Gaston Cremieux, CP 5706, 91057 EVRY cedex, FRANCE. (E-mail :
segref@genoscope.cns.fr - Web : www.genoscope.cns.fr)

COMMENT
This GSS is part of a random genomic sequencing program of thirteen
yeast species: *Saccharomyces bayanus* var. *uvarum*, *Saccharomyces*
exiguus, *Saccharomyces servazzii*, *Zygosaccharomyces rouxii*,
Saccharomyces kluyveri, *Kluyveromyces thermotolerans*, *Kluyveromyces*
lactis var. *lactis*, *Kluyveromyces marxianus* var. *marxianus*, *Pichia*
angusta, *Debaryomyces hansenii* var. *hansenii*, *Pichia sorbicola*,
Candida tropicalis and *Yarrowia lipolytica*. Genomic inserts of 3 to
5 kb were prepared and both extremities were sequenced. See
keywords for description of this sequence and for the sequence of
the other extremity of this insert.

FEATURES
source
location/Qualifiers
1..959
/organism="Saccharomyces kluyveri"
/mol_type="genomic DNA"
/strain="CBS 3082"
/db_xref="taxon:4934"
/clone="AU0A005F10"
/clone_11b="AU0AA"
/note="end : 17"
complement(<71..>958)
/note="similar to *Saccharomyces cerevisiae* ORF YPR186c [PZ1 : TR11a (transcription initiation factor)]"
/evidence=not_experimental

Query Match 13.6%; Score 167.6; DB 9; Length 959;
Best Local Similarity 57.2%; Pred. No. 5.9e-23;
Matches 324; Conservative 1; Mismatches 235; Indels 6; Gaps 1;

Qy 211 CATTGGAAACATATTTGTATCATCTCCGAAAAAACCATTTCATTGTCAGTGT 270

Db 958 CACTTGGAAAGACATATTTGATTCACCGATTCGATATTAACATTTCAATTTGT 899

Qy 271 GGTAAAGGGTTAATTCGCAACACTTGAAGAGACATGAATACCATCAATGAAGCA 330

Db 898 GGTAAAGGTGTACTACAAAGACGATTAATAAGACATGATACATACCAATCG 839

Qy 331 TTTAAATGTACATTTGAAATTTGTCAAGACATTTTAAACATCAATCTTTAAGACAT 390

Db 838 TTCAATGTGAATATGAGGGCTGCAAGAAAGTTTTCAGAGCTTCACATTAAGTCA 779

Qy 391 CATATATATCTGTTTCATGAAAAAATTAACGTTAAACATGATTAAGATTTCAC 450

Db 778 CATACACTTGGTACACTTGCAGAAATGACCTGTGACATTCGGGTAAATGTTCAA 719

Qy 451 CGACCTTGAATTAAGCAACATTAATTAACATCATGATCTCCGCT----- 504

Db 718 AGGCAATATCGTCTGCAAAACCATCTTGCAAAAGCATCAATGTGAGATTTGAACAA 659

Qy 505 TATCATGTGATCATCTGTTGTTTAAATTTCCAACTTGTGTCAGTATTAATTT 564

Db 658 TATCAATGTACTTACAGGGCTGACACAGACTTTTAAACCTGACGCTTTCAGAG 599

Qy 565 CATATATAAACAATGATCAAAAATTAATGTCCTTAATGTGTAAGGTTGTGCG 624

Db 598 CATATCAAGGGGAGCAATCCCAAGTTGCACTGTAACGTATGTGGCAAGCTTGTGTGT 539

Qy 625 AAAAAAGTTATCTTCATATGTTAGTATGATGATGTTCTACATGATCAAAATAGG 684

Db 538 GAGTCTGATCTACAGATGACATGAGTCCATGACGAGGCTTTAGTATTAATAATTTGG 479

Qy 685 ACTTGATTTATTTGATGTGGGAAATTTCCAAAGAAATGAATTTAGTGAACATTT 744

Db 478 AAATGTACATTTTGTGACGAAGTCTCTTTGCTAAGAAAGGCTGATCTTTATCTAC 419

Qy 745 AATATCTTCATGATGATTAATCTCC 770

Db 418 ATGGAACATCATTAAGATGATATACC 393

RESULT 3
CN759437
LOCUS
CN759437 801 bp mRNA linear EST 20-MAY-2004

DEFINITION ID0AAA25CC01RMI ApMS Acyrthosiphon pisum cDNA clone ID0AAA25CC01 5', mRNA sequence.

ACCESSION CN759437

VERSION CN759437.1 GI:47533360

KEYWORDS EST.

SOURCE Acyrthosiphon pisum (pea aphid)

ORGANISM Acyrthosiphon pisum

REFERENCE Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Paraneoptera; Hemiptera; Sternorrhyncha; Aphidiformes; Aphidoidea; Aphididae; Macrosiphini; Acyrthosiphon.

AUTHORS 1 (bases 1 to 801)

Stern,D., Tagu,D. and Wincker,P.

An expressed sequence tags database for the pea aphid Acyrthosiphon pisum

TITLE Unpublished (2004)

JOURNAL Contact: D. Tagu

COMMENT INRA Rennes

UMR Bio3P, BP 35327, F-35653 Le Rheu Cedex France

Tel: +33.2.23.48.51.65

Fax: +33.2.23.48.51.50

Risk of contamination by bacterial sequences from obligatory (Buchnera) or facultative endosymbionts. These sequences were obtained in the frame of the International Consortium of Aphid Genomics in collaboration with Genoscope

PCR Primers

FORWARD: CAGGAAACAGCTATGACC

Plate: 25 row: C column: 1.

FEATURES

source

1..801

Location/Qualifiers

/organism="Acyrthosiphon pisum"

/mol_type="mRNA"

/cultivar="developmentstage"

/db_xref="taxon:7029"

/clone="ID0AAA25CC01"

/tissue_type="whole insect"

/dev_stage="nymphs and adults (parthenogenetic females)"

/lab_host="Xll-Blue"

/clone_lib="ApMS"

/note="Vector: pBS-SK minus; Site_1: EcoRI; Site_2: XhoI; Sample name: ID0AAA ; Plant growth place: Department of Ecology & Evolutionary Biology, Princeton University ; Soil conditions: Soil ; Sowing date: 01/06/1999 ; Harvesting date: 01/06/1999 ; Stress date: no stress ; Description: Aphids inoculated on one-week old Vicia faba under non-sterile conditions. All parthenogenetic stages and both winged and wingless adults were collected for library construction. ; experimental condition: long photoperiod (16-hr light/8-hr dark at 18 c)"

ORIGIN

Query Match 8.4%; Score 103.8; DB 7; Length 801;

Best Local Similarity 51.5%; Pred. No. 2.6e-10;

Matches 318; Conservative 0; Mismatches 287; Indels 12; Gaps 3;

82 ACATATGAAGGGTGTAAAGCCTATATGACCATCATTTATAGCAACATTTAGA 141

8 AAATGATTTCTGTGATAAAGATTTACATATATACAGTTTAAAGCAATCAAG 67

142 ACCACAGTATGATGACCGGTATTAATGATGACGTGATGATTAAGCATTTTC 201

68 ACACATACCGGTGAAAAGCCGTATGATGT-----GATATCTGTGATCAAGCCTTTCT 121

202 AGAAATTCATTTGGAACAACATATTTGATCATTTCCGAAAAAACAATTCATTTGT 261

122 TTGAATTCACATTTAATAGCCATATAATGACATACCGGCGAAAAGCCTTTAATATGT 181

262 TCACTGTGTGTAAGGGTTAATTTTCGACACAACATTGAAAAGACATGAATATCCAT 321

182 GATTAAGTGTATCAAGGGTTTCTTAAGAAATCAAAATTTAACAAGCATATTAAGACAT 241

322 ACAAGTCATTTAATGATGATTTGAAAATTTGTCAAGAGCATTTTATTAACATCAATCT 381

Db 242 ACCGCGAAAAGCCGTATTAATGTGAAAAGTGAACCAAGCGTTTCTCAGAAATCAGAT 301

Qy 382 TTAAAGCATCATATATATCTGTTCAAT---GAAAAACAATTACGTGAACAATGTAAT 438

Db 302 TTAATAGTCATATTAAGACACATACCGGCGAAAAGCGATTAATATGTTAATCTGAT 361

Qy 439 AAAGTTTCACTGACCTTCAAAATTTAGACAACATTAATTAACATCATGTGTGATCT 498

Db 362 CAAGCGTTTCTCAAAAAGTATGTTTAATTAAGCATTAATTAAGCATACCGGTGAAG 421

Qy 499 CTGCTTATCATGTGATCATCTCGTGTGTTTAAAAATTTCCAAACTGTGTAGATTA 558

Db 422 C---CGTTTAAATGTCCTAATCTGTGATCAAGCGTTTATTCGAAAACCATTTAAAAAGC 478

Qy 559 CAATTCATATTAACACATGCATCCAAAACCTTAATGTCCTTAATGTGTTAAAGTTGT 618

Db 479 CACTCAAGACACATACAGCCAAAAGCCGTATTAATGATATCATGTAAACAAGCGTT 538

Qy 619 GTTGGGAAAAAGGTTTATCTTCACATATGTTAAGTCATGATGATTTCAACATGATCA 678

Db 539 TCTCAAAATCAATTTTAACAAGCATTAACAAGACACATGCCGTGAAAAGCTTTAA 598

Qy 679 ATATGACTTGTGATTA 695

Db 599 TGCATTAAGTGTATTA 615

RESULT 4

CN761003 898 bp mRNA linear EST 20-MAY-2004

LOCUS ID0AAA2AE05RMI ApMS Acyrthosiphon pisum cDNA clone ID0AAA2AE05 5', mRNA sequence.

DEFINITION CN761003

ACCESSION CN761003

VERSION CN761003.1 GI:47534926

KEYWORDS EST.

SOURCE Acyrthosiphon pisum (pea aphid)

ORGANISM Acyrthosiphon pisum

REFERENCE Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Paraneoptera; Hemiptera; Sternorrhyncha; Aphidiformes; Aphidoidea; Aphididae; Macrosiphini; Acyrthosiphon.

AUTHORS 1 (bases 1 to 898)

Hunter,W., Martinez-Torres,D., Rahbe,Y., Sabater-Munoz,B., Stern,D., Tagu,D. and Wincker,P.

An expressed sequence tags database for the pea aphid Acyrthosiphon pisum

TITLE Unpublished (2004)

JOURNAL Contact: D. Tagu

COMMENT INRA Rennes

UMR Bio3P, BP 35327, F-35653 Le Rheu Cedex France

Tel: +33.2.23.48.51.65

Fax: +33.2.23.48.51.50

Risk of contamination by bacterial sequences from obligatory (Buchnera) or facultative endosymbionts. These sequences were obtained in the frame of the International Consortium of Aphid Genomics in collaboration with Genoscope

PCR Primers

FORWARD: CAGGAAACAGCTATGACC

Plate: 2 row: E column: 5.

FEATURES

source

1..898

Location/Qualifiers

/organism="Acyrthosiphon pisum"

/mol_type="mRNA"

/cultivar="developmentstage"

/db_xref="taxon:7029"

/clone="ID0AAA2AE05"

/tissue_type="whole insect"

/dev_stage="nymphs and adults (parthenogenetic females)"

/lab_host="Xll-Blue"

/clone_lib="ApMS"

/note="Vector: pBS-SK minus; Site_1: EcoRI; Site_2: XhoI; Sample name: ID0AAA ; Plant growth place: Department of Ecology & Evolutionary Biology, Princeton University ; Soil conditions: Soil ; Sowing date: 01/06/1999 ;

ORIGIN

Harvesting date: 01/06/1999; Stress date: no stress; Insect rearing: Aphids inoculated on one-week old *Vicia faba* under non-sterile conditions. All parthenogenetic stages and both winged and wingless adults were collected for library construction; experimental condition: long photoperiod (16-hr light/8-hr dark at 18 °C)

/tissue_type="olfactory brain"
/clone_lib="RIKEN full-length enriched mouse cDNA library"
/dev_etage="adult"
330..1223
/note="unnamed protein product; MSZF3 (FRAGMENT) homolog
[mus musculus] (SPR|088251, evidence: FASTA, 87.1%id,
100%length, match=255)
putative"

/codon_start=1
/protein_id="BAC27766.1"
/db_xref="GI:26328051"
/translation="MLETYRNALNVAAGYNEDSNIEHCSSRRHGRHNRHTEKPYE
GIQYGBAPVHSSILQMKIITHGEKRYKNCQCKAYSRHSILQIHKRTSGEPYCN
OGKAPTOHSHLKIHWHTGKPYKNCQCKAPAHNLQKERTHTGKPYKQDCC
NKAFTVSYLIQVHKHTHTGKPYKNCQCKAPRHSILQIHKRTSGEPYKNCQCK
ALAHSTLQVHHTGKPYKNCQCKAPFANSTYQVHNRHTGKPYKQDCKGAF
VSSSDLRHHRVHTGKPYKNCQCKAPFANSTYQVHNRHTGKPYKQDCKGAF
1594..1599
/note="putative"
/note="putative"
1609
/note="putative"

Query Match 7.9%; Score 98.2; DB 3; Length 1609;
Best Local Similarity 51.2%; Pred. No. 3.6e-09;
Matches 310; Conservative 0; Mismatches 283; Indels 12; Gaps 3;

60 ACCTCCCAAAAGTATATTTGACATATGAAAGGTGATTAAGCCTTAATGACACATC 119
DB ACATCTGAGAGAAACGCTACAAATGCAATGATGATTAAGCCTTAATGACACAG 580
QY 120 ATTATAGACACATTGAACCCACAGTATGATCGACCTTAATTAATGATCAGTGA 179
DB 581 CATTCTACAAATCATTAAGAAACACATGATGAGAAACCTTAATGAATGA-----A 634
QY 180 CGATTGTGATTAAGCAATTTTTCAGAAATATCAATTTGAAACACATATTTGATCATTC 239
DB 635 TCATGTGTGATTAAGCCTTTACACACACATCATCTCAAAATACATATGTTACATAC 634
QY 240 CGAAAAAAACCATTCATTTGATGATGATGATGATGATGATGATGATGATGATGAT 299
DB 695 TGAGAGAAACCTTACAAATGATGATGATGATGATGATGATGATGATGATGATGAT 754
QY 300 GAAAAACATGAATATCAACCATCAATGATGATGATGATGATGATGATGATGATGATG 359
DB 755 TCAAAAAACATGAAGAAACATCTGAGAGAAACCTTACAAATGATGATGATGATGAT 814
QY 360 AGCATTT---TATTAACATCAATCTTAAAGACATCATATTTATCTGTTGATGAAAAAC 416
DB 815 AGCCTTTGATTAAGAAAGTTATTTACAGATTTCAAAAAACACATATCTGAGAGAAAC 874
QY 417 ATTACGTTTAAACAATGATTAAGTTTCACTCGACCTTCAAAATTAGACACATTA 476
DB 875 TTTCAAATGTAAGAAATGTTAAAGCCTTTGACACGACACATCATCTCAAAATGATGA 934
QY 477 ATTAAAAACATCATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 536
DB 935 AATTAACACATATCTGAGAGAAAC---CTTACAAATGTAAGAAATGTTAAAGCCTTGC 991
QY 537 TTTCCAAACCTTGGTCAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 596
DB 992 ATTATCATATGATCACTCAAGTACATCAAAACACATCTGAGAGAAAGCCTTAATGAT 1051
QY 597 TCCTAAATGTGTAAGAGTTGTGTGGAAGAAAGTTTATCTTCAATATGTTAAATGCA 656
DB 1052 TGAGCAATGTGCAAGCCTTTGCAATCAAAAGTTATTTCAAGTACATTAAGAAATACA 1111
QY 657 TGATG 661
DB 1112 TACTG 1116

BC047646
LOCUS BC047646 2647 bp mRNA linear HTC 19-NOV-2003
DEFINITION Homo sapiens zinc finger protein 85 (HPF4, HTF1), mRNA (cDNA clone
IMAGE:4828290), with apparent retained intron.
ACCESSION BC047646
VERSION BC047646.1 GI:29126804
KEYWORDS HTC.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens

REFERENCE
AUTHORS Klausner, R.D., Collins, F.S., Wagner, L., Shemen, C.M., Schlegel, G.D.,
Altechul, S.F., Zeeberg, B., Burow, K.H., Schefer, C.F., Bhat, N.K.,
Hopkins, R.F., Jordan, H., Moore, T., Max, S.I., Wang, J., Hsieh, F.,
Diachenko, L., Marusina, K., Farmer, A.A., Rubin, G.M., Hong, L.,
Stapleton, M., Soares, M.B., Bonaldo, M.F., Casavant, T.L.,
Scheetz, T.E., Brownstein, M.J., Usdin, T.B., Toshiyuki, S.,
Carninci, P., Prange, C., Raha, S., Loquellano, N.A., Peters, G.J.,
Abramson, R.D., Mullany, S.J., Bosak, S.A., McEwan, P.J.,
McKernan, K.J., Malek, J.A., Gunaratne, P.H., Richard, S.,
Worley, K.C., Hale, S., Garcia, A.M., Gay, L.J., Hulyk, S.W.,
Villalón, D.K., Muzny, D.M., Sodergren, E.J., Lu, X., Gibbs, R.A.,
Fahey, J., Helton, E., Kettelman, M., Madan, A., Rodriguez, S.,
Sanchez, A., Whiting, M., Madan, A., Young, A.C., Shevchenko, Y.,
Bouffard, G.G., Blakeley, R.W., Touchman, J.W., Green, E.D.,
Dickson, M.C., Rodriguez, A.C., Grimwood, J., Schultz, J., Myers, R.M.,
Butterfield, Y.S., Krzywinski, M.I., Skelton, D., Small, D.B.,
Scherer, A., Schein, J.E., Jones, S.J., and Marra, M.A.

Generation and initial analysis of more than 15,000 full-length
human and mouse cDNA sequences
Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)
22388257
MEDLINE 12477932
JOURNAL 2 (bases 1 to 2647)
PUBMED 12477932
REFERENCE Strausberg, R.
AUTHORS Direct Submission
TITLE Submitted (03-MAR-2003) National Institutes of Health, Mammalian
Gene Collection (MGC), Cancer Genomics Office, National Cancer
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
USA

REMARK NIH-MGC Project URL: <http://mgc.ncl.nih.gov>
COMMENT Contact: MGC help desk
Email: cgabbs-remail.nih.gov
Tissue Procurement: Miklos Palokovits, M.D., Ph.D.
CDNA Library Preparation: Michael J. Brownstein (NHGRI) & Shiraki
Toshiyuki and Piero Carninci (RIKEN)
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: Sequencing Group at the Stanford Human Genome
Center, Stanford University School of Medicine, Stanford, CA 94305
Web site: <http://www-shgc.stanford.edu>
Contact: (Dickson, Mark) medpax1.stanford.edu
Dickson, M., Schmutz, J., Grimwood, J., Rodriguez, A., and Myers,
R. M.

Clone distribution: MGC clone distribution information can be found
through the I.M.A.G.E. Consortium/LNL at: <http://image.llnl.gov>
Series: IRAP Plate: 90 Row: 1 Column: 7
This clone has the following problem: retained intron.
Location/Qualifiers
1..2647
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:4828290"
/issue_type="Testis"
/clone_lib="NIH_MGC_97"
/lab_host="DH10B"
/note="Vector: pBluescript"

RESULT 6

Query Match

7.9%; Score 97.8; DB 3; Length 2647;

Best Local Similarity 50.1%; Pred. No. 4.4e-09;
Matches 330; Conservative 0; Mismatches 317; Indels 12; Gaps 3;

```

QY 7 GAAAGTACGAAACCAATGATATATCTTTTAATATCTTCTTCTTCATCAGCTCC 66
    |||||
Db 1235 GAATGTGGCAAACTTTTAAACGATCTCTCACTCTACCAATGAATATCTACT 1294
    |||||
QY 67 AAAAGTATATTTTGCATATGAAAGGTGTGATTAAGCCTATTAATGACCATTATTA 126
    |||||
Db 1295 GGAGAGAAACCCCTACAAATGTAAAGATGTGTAAGCCTTTTAACGATCTCAACCTT 1354
    |||||
QY 127 GAGCAACATTTAAGAACCCAGTAAATGATCCGCTATTAATGATACAGGCGATTTGT 186
    |||||
Db 1355 ACTACCAATGAAAAATTCATCTGAGAGAAACCTTACAAAGT-----GAAAGATGT 1408
    |||||
QY 187 GATAAAGCATTTTTCAGAAAATCACATTTGGAACAATATTTGATCATCTCCGAAAA 246
    |||||
Db 1409 GCGAAAGCCTTTAAGAGAGTCTCAAACTTACATCACTAAGATTAATCTACTGAGAG 1468
    |||||
QY 247 AAACATTCATTTTCACTGTGTGTGTGTAAGGGGTAAATTTCTGACAACTTGAAAA 306
    |||||
Db 1469 AAACCTACAAATGTAAAAAATGTGAAAAAGCCTTTAACAGTGTGACACTTACACA 1528
    |||||
QY 307 CATGAATACCCATACAAAGTATTTAATGTATCATTTGAAATTTGTCAAGAGCATTT 366
    |||||
Db 1529 CATGAGTATTTATCTGAGAGAAACCCCTACAAATGTGAAAAATGTGAAAAAGCCTTT 1588
    |||||
QY 367 TATTA---ACATCAATCTTTAAGACATCATATATATTTCTGATGAAAAAACATTAA 423
    |||||
Db 1589 AATCATTTCTCACACCTTACTACATAGATATATTCATATCTGAGAGAAACCTTACAA 1648
    |||||
QY 424 TGTAAACAATGTAAAGTTTCACTCGACCTTCAAAATTAAGACACATTAATTA 483
    |||||
Db 1649 TGTAAAGATGTGTAAAGCTTTTAAACACTCTTCAACCTTACTAAACATTAAGATTA 1708
    |||||
QY 484 CATCATGTGTGATCTCTGCTTATCATGTGATCATCTGCTGTTTAAAAATTTCGA 543
    |||||
Db 1709 CATACGTGAGAGAAAGCTTACAAATGTAAAGATGTGAAAAAGCCTTTTAAACATCTCA 1768
    |||||
QY 544 ACTGTGTCATTTTATTAATTTTCAATTAATAACAATGCTCAAACTTAAATGCTTAA 603
    |||||
Db 1769 AAACCTACTGACATTAAGAAAAATTCATCTGAGAGAAACCTTA---TGAATGTAAAA 1825
    |||||
QY 604 TGTGTAAAGTTGTGTGGAAAAAAGTTTATCTTCAATATGTTAATGATGATGA 662
    |||||
Db 1826 TGTGCAAAAGCTTTTAAACGAGTCTCAAAATCTTACTAGACATTAAGAAAAAGTCA 1884
    |||||

```

RESULT 7
AUI23448 745 bp mRNA linear EST 01-AUG-2002
LOCUS AUI23448 NT2RM2 Homo sapiens cDNA clone NT2RM2000319 5', mRNA
DEFINITION sequence.

ACCESSION AUI23448
VERSION AUI23448
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens

REFERENCE
AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 745)
Ota, T., Wakamatsu, A., Ozawa, M., Ishii, S., Saito, K., Yamamoto, J.,
Nakamura, Y., Nishikawa, T., Nagai, T., Suzuki, Y., Sugano, S. and
Isogai, T.

TITLE HRI human cDNA project (Ota, T., Wakamatsu, A., Ozawa, M., Ishii, S.,
Saito, K., Yamamoto, J., Nakamura, Y., Nishikawa, T., Nagai, T.,
Suzuki, Y., Sugano, S., Isogai, T.)
Unpublished (2000)
Contact: Takao Isogai
Genomics Laboratory
Helix Research Institute
1532-3 Yana, Kisarazu, Chiba 292-0812, Japan
Tel: 81-438-52-3975

Fax: 81-438-52-3986
Email: genomics@hri.co.jp
HRI human cDNA project; 5'- & 3'-end one pass sequencing; Helix
Research Institute; cDNA library construction; Department of
Virology, Institute of Medical Science, University of Tokyo, and
Helix Research Institute.

FEATURES

source
1..745
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="NT2RM2000319"
/cell_type="teratocarcinoma"
/cell_line="NT2"
/clone_1ib="NT2RM2"
/note="Vector: pME18PFL3; mRNA from uninduced NT2 neuronal
precursor cells"

ORIGIN

Query Match 7.9%; Score 97.6; DB 1; Length 745;
Best Local Similarity 50.2%; Pred. No. 4.4e-09;
Matches 328; Conservative 0; Mismatches 314; Indels 12; Gaps 3;

```

QY 7 GAAAGTACGAAACCAATGATATCTTTAATATCTTCTTCTTCATCAGCTCC 66
    |||||
Db 87 GAATGTGGCAAACTTTTAAACGATCTCTCACTCTACCAATGAATATCTACT 146
    |||||
QY 67 AAAAGTATATTTTGCATATGAAAGGTGTGATTAAGCCTATTAATGACCATTATTA 126
    |||||
Db 147 GAGAGAAACCCCTACAAATGTAAAGATGTGTAAGCCTTTTAAACGCTTCTTCAACCTT 206
    |||||
QY 127 GAGCAACATTTAAGAACCCAGTAAATGATCCGCTATTAATGATACAGGCGATTTGT 186
    |||||
Db 207 ACTACCAATGAAAAATTCATCTGAGAGAAACCTTCAAAAGT-----GAAAGATGT 260
    |||||
QY 187 GATAAAGCATTTTTCAGAAAATCACATTTGGAACAATATTTGATCATCTCCGAAAA 246
    |||||
Db 261 GCGAAAGCCTTTAAGAGAGTCTCAAACTTACTACATTAAGATATTCATCTGAGAG 320
    |||||
QY 247 AAACATTCATTTTCACTGTGTGTGTGTAAGGGGTAAATTTCTGACAACTTGAAAA 306
    |||||
Db 321 AAACCTACAAATGTAAAAAATGTGAAAAAGCCTTTAACAGTGTGACACTTACTACA 380
    |||||
QY 307 CATGAATACCCATACAAAGTATTTAATGTATCATTTGAAATTTGTCAAGAGCATTT 366
    |||||
Db 381 CATGAGTATTTATCATCTGAGAGAAACCCCTACAAATGTGAAAAATGTGAAAAAGCCTTT 440
    |||||
QY 367 TATTA---ACATCAATCTTTAAGACATCATATATTTCTGATGAAAAAACATTAA 423
    |||||
Db 441 AATCATTTTCTCACACCTTACTACATTAAGATTAATTCATCTGAGAGAAACCTTACAA 500
    |||||
QY 424 TGTAAACAATGTAAAGTTTTCATGACCTTCAAAATTAAGACACATTAATTA 483
    |||||
Db 501 TGTAAAGATGTGTAAAGCTTTTAAACCTTCAACCTTACTTAACATTAAGATTA 560
    |||||
QY 484 CATCATGTGTGATCTCTGCTTATCATATGATCATCTGCTGTTTAAAAATTTCGA 543
    |||||
Db 561 CATCTGAGAGAAACCTTACAAATGTAAAGATGTGAAAAAGCTTTTAAACCAATCTCA 620
    |||||
QY 544 ACTGTGTCATTTTATTAATTTTCAATTAATAACAATGCTCAAACTTAAATGCTTAA 603
    |||||
Db 621 AAACCTTACTGAAACATTAAGAAATTCATCTGAGAGAAACCTTA---TGAATGTAAAA 677
    |||||
QY 604 TGTGTAAAGTTGTGTGGAAAAAAGTTTATCTTCAATGTTAATGATAT 657
    |||||
Db 678 TGTGCAAAAGCTTTTAAACGAGTCTCAAAATCTTACTAGACATTAAGAAAAAGTCA 731
    |||||

```

RESULT 8
CO434402 727 bp mRNA linear EST 06-JUL-2004
LOCUS CO434402
DEFINITION UT-M-HX0-csl-b-09-0-UT-r1 NIH_BMAP_HX0 Mus musculus cDNA clone
IMAGE:30687440 5', mRNA sequence.

ACCESSION CO434402
VERSION CO434402.1 GI:49680696
KEYWORDS EST.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 727)
TITLE NIH-MGC http://mgc.nci.nih.gov/
AUTHORS National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgabs-remail.nih.gov
Tissue Procurement: Dr. James L. University of Iowa
CDNA Library Preparation: Dr. M. Bento Soares, University of Iowa
CDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
Clone Distribution: Distribution information can be found at
http://genome.iowa.edu/distribution/mousefl.html
This clone was contributed by the Brain Molecular Anatomy Project (BMAP)
Seq primer: pyx-5.
FEATURES
source
Location/Qualifiers
1..727
/organism="Mus musculus"
/mol_type="mRNA"
/db_xref="taxon:10090"
/clone="IMAGE:30687440"
/rname_type="whole eye"
/dev_stage="newborn (1, 5, 15 days) and embryonic (15, 16, 17, 18 dpc)"
/lab_host="DH10B (T1 phage resistant)"
/clone_id="N1H_BMAP_HK0"
/note="Organ: Eye; Vector: pyx-Asc; Site 1: EcoR I; Site 2: Not I; The library was constructed according to Bonaldo, Lennon and Soares, Genome Research, 6:791-806, 1996. Denatured RNA was size fractionated on a 1% agarose gel. First strand cDNA synthesis was primed with oligo-dT primer containing a Not I site. Double strand cDNA was size selected according to mRNA size fraction, ligated with EcoR I adaptor, digested with NotI and then cloned directionally into pyx-Asc vector. The library tag sequence located between the Not I site and the polyA tail is AATTAATACG. This library was created for the University Iowa Brain Anatomy Project (BMAP): 'Gene Discovery in the Developing Mouse Nervous System', supported by National Institute of Mental Health (NIMH)."

ORIGIN
Query Match 7.6%; Score 93.4; DB 7; Length 727;
Best Local Similarity 52.1%; Pred. No. 3e-08;
Matches 261; Conservative 0; Mismatches 231; Indels 9; Gaps 2;
4 AGTGAAGTACGAAACCAATGATATCATCTTATATCTTCTTCATCAAGT 63
Db AATCAGTGTATTAAGCCTTTCTCGATACGATCTTCAATCTCAATGAAGACAT 137
64 CCCAAAAGTATATTTGCACATATGAAGGTGTATTAAGCCTATATTCACCATCTTA 123
Db ACTGAGAGAAACCTTCAATGTATCAAGTGTATTAAGCCTTTTGTGAAGATGAGT 197
124 TTGAGCAACATTTAAAGAACCCACAGTATGATGACCGTATTAATGTATACAGTGAGCAT 183
Db CTCGAAAACCATGTAAAGATACATACCTGAGAGAAACCTTACAAATGTA-----ATCAA 251
184 TGTGATAAGCATTTTCAGAAATTCACATTTGGAACACATATTTGATATCATTTCCGAA 243
Db TGTGATTAAGCCTTTTCCCATATACCTCCACATTTATTAAGAGAACATACCGGA 311
244 AAAAACCATTTCATTTGTGAGTGTGTAAAGGGTTAATTTCTGCACAACTTGAA 303
Db GAGAAACCTTAACAAATGTATCAAGTGTATTAAGCCTTTCCCATATCTATCAGCTCCAC 371

ORIGIN
Query Match 7.5%; Score 92.4; DB 7; Length 631;
Best Local Similarity 54.1%; Pred. No. 4.7e-08;
Matches 236; Conservative 0; Mismatches 191; Indels 9; Gaps 2;
304 AGACATGAATCACCACATCAAAAGTCATTTAAATGTACATTTGAATTTGTCAAGAGCA 363
Db ATTCAATGAAGAACACATCTGAGAGAAAGCCTTACAAATGTATATGATGTGATTAAGCC 431
364 TTTTATTAACATCAATCTTTTAAGACATCATATATATTC--TGTGATGAAAAACATTA 420
Db TTTTTCATATCTTTCTCTTCAAACTACAGAAAGAACATATCTGAGAGAAACATTC 491
421 ACCTGTAAACATGTATTAAGTTTCACTCGACCTTCAAAATATGACACAACTTAATTA 480
Db AATGTAAACAAATGTATTAAGCTTTTCTCATATACGATATCTTCAAACTCATAGAGA 551
481 AAACATCATGTGATCTCCT 501
Db AACATATCTGAGAGAAACCT 572

RESULT 9
CN352904
LOCUS CN352904 631 bp mRNA linear EST 16-MAY-2004
DEFINITION 17000600057466 GRN_PREHEP Homo sapiens cDNA 5', mRNA sequence.
ACCESSION CN352904
VERSION CN352904.1 GI:47352838
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 631)
AUTHORS Brandenberger R., Wei, H., Zhang, S., Lei, S., Muraie, J., Flak, G. J., Li, Y., Xu, C., Fang, R., Guegler, K., Rao, M. S., Mandalam, R., Lebkowski, J. and Stanton, L. W.
TITLE Transcriptional characterization elucidates signaling networks that control human ES cell growth and differentiation
JOURNAL Nat. Biotechnol. 22 (6), 707-716 (2004)
COMMENT Contact: Brandenberger R
Regenerative Medicine
Geron Corporation
230 Constitution Drive, Menlo Park, CA 94025, USA
Tel: 650 473 8658
Fax: 650 473 7760
Email: rbrandenberger@geron.com
Insert Length: 631 Std Error: 0.00.
Location/Qualifiers
1..631
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/rname_type="embryonic stem cells, DMSO-treated H9 cell line"
/note="oligo dT primed, full-length enriched cDNA library from DMSO-treated hES cell line H9 (p22) maintained in feeder-free conditions"

ORIGIN
Query Match 7.5%; Score 92.4; DB 7; Length 631;
Best Local Similarity 54.1%; Pred. No. 4.7e-08;
Matches 236; Conservative 0; Mismatches 191; Indels 9; Gaps 2;
69 AAGTATATTTGACATATGAAGGTGTATTAAGCCTATATTCAGACATATTAATTAAGA 128
Db AGAGAAAGCCTTAAATGTAAAGATGTAAAAAGCTTTTAAACCAATCTCCAAACCTTAC 142
129 GCAACATTTAAAGAACCCACAGTATGATGACCGTATTAATGTATACAGTGAGCATTTGTGA 188
Db TGAACATTAAGAAATTCATATCTGAGAGAAACCTTATGATGT-----GAAAAATGTGG 196
189 TAAAGCATTTTTCAGAAAAATTCACATTTGGAACACATATTTGATATCATTTCCGAAAAAA 248
Db CAAAGCTTTTAAACCAAGTCTCAATCTTACTAGACATTAAGAAAAAGTCATACAGAGAGAA 256

QY 249 ACCATTCATTGTCAGTGTGTGTAAGGGGTAAATCTCGACAACCTTGAAAGACA 308
DB 257 ACCCTTCAATGTGAAGATGTGGCAAGGTTTAAATGAGCCCTCAACCTTACTATCA 316
QY 309 TGAATACCCATCAAGATGATTTAAATGATCTTGAATATGTCAGAAAGCATTTTA 368
DB 317 TAAAGATTAATTCATACGTGAGAGAAACCATTAATGTAAGATGTGGCAAGCTTTTA 376
QY 369 TAAA---CATCATCTTTTAAGACATCATATATATCTGTTCAATGAATAAATTAACGTG 425
DB 377 CCATTCCTCAAAACTTACCAACATTAAGAAATTCATCTAGAGAGAAACCTTACACATG 436
QY 426 TAAACATGTATTAAGATTTCATCTGACCTTCAAAAATTAGCAACAATTAATTAACA 485
DB 437 TGAAGATGTGGCAAGGCTTTAACAGTCTCAAACTTACTAATCAATTAAGAAATTC 496
QY 486 TCATGTGTGATCTCT 501
DB 497 TACTGGAGAAAAAACCCT 512

RESULT 10
BC036394 2672 bp mRNA linear HTC 19-NOV-2003
LOCUS
DEFINITION Homo sapiens zinc finger protein 85 (HPF4, HTP1), mRNA (cDNA clone IMAGE:5259399), containing frame-shift errors.
ACCESSION BC036394
VERSION BC036394.1 GI:23025784
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
1 (bases 1 to 2672)
Straussberg, R.L., Feingold, E.A., Grouse, L.H., Derge, J.G., Klausner, R.D., Collins, F.S., Wagner, L., Shenmen, C.M., Schuler, G.D., Altshuler, S.P., Zeeberg, B., Buelow, K.H., Scheffer, C.F., Bhat, N.K., Hopkin, R.F., Jordan, H., Moore, T., Max, S.I., Wang, J., Hsieh, F., Diachenko, L., Marusina, K., Farmer, A.A., Rubin, G.M., Hong, L., Stapleton, M., Soares, M.B., Bonaldo, M.F., Casavant, T.L., Scheetz, T.E., Brownstein, M.J., Uedin, T.B., Toshiyuki, S., Carninci, P., Prange, C., Raha, S.S., Loquellano, N.A., Peters, G.J., Abramson, R.D., Mullaly, S.J., Bosak, S.A., McEwan, P.J., McKernan, K.J., Malek, J.A., Gunaratne, P.H., Richards, S., Wootley, K.C., Hale, S., Garcia, A.M., Gay, L.J., Hult, S.W., Villalón, D.K., Muzny, D.M., Sodergren, E.J., Lu, X., Gibbs, R.A., Fahey, J., Helton, E., Kettelman, M., Madan, A., Rodriguez, S., Sanchez, A., Whiting, M., Madan, A., Young, A.C., Schevchenko, Y., Bouffard, G.G., Blakeley, R.W., Touchman, J.W., Green, E.D., Dickson, M.C., Rodriguez, A.C., Grimwood, J., Schmutz, J., Myers, R.M., Butlerfield, Y.S., Krzywinski, M.I., Skalska, U., Small, D.E., Scherch, A., Schein, J.E., Jones, S.J. and Marra, M.A.
Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences
Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)
2 (bases 1 to 2672)
12477932
22388257
PUBMED
JOURNAL
MEDLINE
REFERENCE
AUTHORS
TITLE
JOURNAL
REMARK
COMMENT
NHL-MGC Project URL: <http://mgc.nci.nih.gov>
Contact: MGC help desk
Email: cgabbs-r@mail.nih.gov
Tissue Procurement: Miklos Palkovits, M.D., Ph.D.,
CDNA Library Preparation: Michael J. Bronstein (NHGRI) & Shizaki
Toshiyuki and Piero Carninci (RIKEN)
CDNA Library Arrayed by: The I.W.A.G.E. Consortium (LINL)
DNA Sequencing by: Institute for Systems Biology
<http://www.systemsbio.org>

contact: amadan@systemsbio.org
Anup Madan, Jessica Fahey, Erin Helton, Mark Kettelman, Anuradha Madan, Stephanie Rodrigues, Amy Sanchez and Michelle Whiting
Clone distribution: MGC clone distribution information can be found through the I.W.A.G.E. Consortium/LINL at: <http://image.linl.gov>
Series: IRAP Plate: 73 Row: a Column: 16
This clone has the following problem: frame shifted.
Location/Qualifiers
1..2672
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:5259399"
/issue_type="Brain, hippocampus"
/clone_id="NHL MGC_95"
/lab_host="DH10B"
/note="Vector: pBluescript"

ORIGIN
Query Match 7.5%; Score 92.4; DB 3; Length 2672;
Best Local Similarity 54.1%; Pred. No. 5,2e-08;
Matches 236; Conservative 0; Mismatches 191; Indels 9; Gaps 2;
QY 69 AAGATATATTTGCACATATGAAGGGTGTGATTAAGCCATATATCGACCATCTATTATGA 128
DB 1745 AGAGAAAGCTTACAAATGTAAAGATGTGAAAAGCTTTTAACCAATCCTCAAACTTAC 1804
QY 129 GCAATTTTAAGAACCCACAGTATGATCGACCGTATTAATGTATACATGACGATTTGA 188
DB 1805 TGACATATGAATAATTCATCTGAGAGAAACCCATATATGT-----GAAAATGTGG 1858
QY 189 TAAACATTTTTCAGAAATCATTTGGAACAATTTGATTCATTCGATTCGAAAAAA 248
DB 1859 CAACCTTTTAACCAAGCTCTCAATCTTACATTAAGCAATTAATCTTACGAAAGAA 1918
QY 249 ACCATTCATTGTCAGTGTGTGTAAGGGGTAAATCTCGACAACAATTGAAAAACA 308
DB 1919 ACCTTACAAAGTGAAGATGTGGCAAGGTTTAAATGTGGCCCTCAACCTTACTATCA 1978
QY 309 TGAATACCCATCAAGATGATTTAAATGTATATTGTAATTTGTAAGAAAGCATTTTA 368
DB 1979 TAAAGTATATTCATCTGAGAGAAACCATTAAGTAAAGATGTGCAAGCTTTTAA 2038
QY 369 TAAA---CATCATCTTTTAAGACATCATATATATCTGTTCAATGAATAAATTAACGTG 425
DB 2039 CCATTCCTCAAAACTTACCAACATTAAGAAATTCATCTAGAGAGAAACCTTACACATG 2098
QY 426 TAAACATGTATTAAGATTTCATCTGACCTTCAAAAATTAGCAACAATTAATTAACA 485
DB 2099 TGAAGATGTGGCAAGGCTTTAACAGTCTCAAACTTACTAATCAATTAAGAAATTC 2158
QY 486 TCATGTGTGATCTCT 501
DB 2159 TACTGGAGAAAAAACCCT 2174

RESULT 11
A0005136 591 bp DNA linear GSS 27-JUN-1998
LOCUS
DEFINITION CIT-HSP-2290017.TP CIT-HSP Homo sapiens genomic clone 2290017,
genomic survey sequence.
ACCESSION A0005136
VERSION A0005136.1 GI:3082581
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
1 (bases 1 to 591)
Adams, M.D., Rounisley, S.D., Zhao, S., Field, C.E., Baas, S., Linher, K., Golden, K., Berry, K., Granger, D., Suh, E., Wible, C., Shizuya, H., Simon, M. and Venter, J.C.

TITLE Use of a random BAC End Sequence Database for Sequence-Ready Map Building (1998)
JOURNAL Unpublished (1998)
COMMENT Other GSBS: CIT-HSP-2290017.TR
Contact: Mark Adams
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: mdamas@tigr.org
Clones are available from Research Genetics (info@resgen.com). BAC end search page:
http://www.tigr.org/tdb/humgen/bac_end_search/bac_end_search.html
Seq primer: M13-21;
Class: BAC ends.

FEATURES
Source Location/Qualifiers
1..591
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="GDB:7150844"
/db_xref="taxon:9606"
/clone="2290017"
/sex="Male"
/cell_type="Sperm"
/clone_1lb="CIT-HSP"
/note="Vector: pBelobAC11; Site_1: HindIII; Site_2: HindIII"

ORIGIN
Query Match 7.4%; Score 92; DB 8; Length 591;
Best Local Similarity 52.2%; Pred. No. 5.6e-08;
Matches 256; Conservative 0; Mismatches 225; Indels 9; Gaps 2;
2;
QY 7 GAAAGTGCAGAAACCAATGCATATCATCTTAAATCTCTCTCTTCATCAGCTCC 66
DB GAATGTGGCAAGCCTTCTACCATCTTTCACACCTTACACATAGATTAATTCATCT 128
QY 67 AAAAGTATTTTGACATATGAAAGGTGTGATTAAGCCATATGACCATCATATTA 126
DB 129 GGAGAGAGGCCCTTCAATGTGAAGATGTGTAAAGCTTTTAAACCCCTTACGCCCT 188
QY 127 GAGCAACATTTAAGAACCCACAGATATGATCGACCGTATTAATGTACAGTGCATTTG 186
DB 189 ACTACACATTAAGTTCATTCATGTTAAAGAAAACCCTCACAAATGT-----GAAATGT 242
QY 187 GATTAAGCATTTTTCAGAAAATCACTTTGAAAACATATTTGATCACTTCGAAAA 246
DB 243 GACAAAGCTTTTAAACGATTCCTATACCTTAAGATTAAGATTAATTCCTGGAGAG 302
QY 247 AAACCATTCATTTGATGTGTGTGTGTAAAGGGGTAAATTCGACAACCTTGAAGA 306
DB 303 AAATCTTCAAAATGTGAACATGTGGCAAGGCTTTAACTGTCTTCACACCTTACAAA 362
QY 307 CATGAATTCACCAATCAAAAGTCATTTAAATGTACATTTGAAAATTTGCAAGAGATT 366
DB 363 CATTAAGAAATTTCTACTGAGAGAAACCTCAAAATGTAAAGATGTGGCAAGCCTTT 422
QY 367 TATTAACAT--CAATCTTTAAGACATCATATATATCTGTGCATGAAAAACATTACG 423
DB 423 AATGTGCTTCAACCTTACTACACATPAAGATGATTCATCTGGAAGAAAACCTTACAA 482
QY 424 TGTAAACATGTATTAAGTTTTCATCTGACCTTCAAAATTTAGACAACATTAATTAAA 483
DB 483 TGTGAAGAAATGTGGCAAGCCTTTAACTCCTCAAACTTACTATATACATAAGATTAAT 542
QY 484 CATCATGTG 493
DB 543 CATCTGGAG 552

LOCUS CN362485 687 bp mRNA linear EST 16-MAY-2004
DEFINITION 17000424188118 GRN_EB Homo sapiens cDNA 5', mRNA sequence.
ACCESSION CN362485
VERSION CN362485.1 GI:47362419
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
REFERENCE 1 (bases 1 to 687)
Brandenberger, R., Wei, H., Zhang, S., Lei, S., Muzage, J., Fisk, G. J., Li, Y., Xu, C., Fang, R., Guegler, K., Rao, M. S., Mandalam, R., Lebkowski, J. and Stanton, L. W.
Transcriptome characterization elucidates signaling networks that control human ES cell growth and differentiation
Nat. Biotechnol. 22 (6), 707-716 (2004)
CONTACT: Brandenberger R
Regenerative Medicine
Geron Corporation
230 Constitution Drive, Menlo Park, CA 94025, USA
Tel: 650 473 8658
Fax: 650 473 7760
Email: rbrandenberger@geron.com
Insert length: 687 Std Error: 0.00.
Location/Qualifiers
1..687
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/feature_type="embryonic stem cells, embryoid bodies derived from H1, H7 and H9 cells"
/clone_1lb="GRN_EB"
/note="ToIigo dT primed, full-length enriched cDNA library from embryoid body outgrowth derived from hES cell lines H1 (p32), H7 (p29), and H9 (p26) maintained in feeder-free conditions."

ORIGIN
Query Match 7.4%; Score 92; DB 7; Length 687;
Best Local Similarity 52.2%; Pred. No. 5.6e-08;
Matches 256; Conservative 0; Mismatches 225; Indels 9; Gaps 2;
2;
QY 7 GAAAGTGCAGAAACCAATGCATATCATCTTAAATCTCTCTCTTCATCAGCTCC 66
DB 183 GAATGTGGCAAGCCTTCTACCATCTTTCACACCTTACACATAGATTAATTCATCT 242
QY 67 AAAAGTATTTTGACATATGAAAGGTGTGATTAAGCCATATGACCATCATATTA 126
DB 243 AGATGAATTTTTCACAAATGTGAAGATGTGAAGAGCCTTTAACTGTCTCAACCTT 302
QY 127 GAGCAACATTTAAGAACCCACAGTATGATCGACCGTATTAATGTACAGTGCATTTG 186
DB 303 AAATCTTCAAAATGTGAACATGTGGCAAGGCTTTAACTGTCTTCACACCTTACAAA 362
QY 187 GATTAAGCATTTTTCAGAAAATCACTTTGAAAACATATTTGATCACTTCGAAAA 246
DB 357 GGCAAGGCCCTTACACGACCTTACACCTTATTAAGCATTAAGATTTCTACTGAAG 416
QY 247 AAACCATTCATTTGATGTGTGTGTGTAAAGGGGTAAATTCGACAACCTTGAAGA 306
DB 417 AAACCTTAAGAAATGTGAAGATGTGGCAAGCCTTTAACTGATCATGACCTTACTACA 476
QY 307 CATGAATTCACCAATCAAAAGTCATTTAAATGTACATTTGAAAATTTGCAAGAGATT 366
DB 477 CATTAATTAATTCATCTGAGGAAATTCCTCAAAATGTGAAGAAATGTGTAGAGCTTT 536
QY 367 TATTAACATCAATCTTTAAGACATCATATATATCTGT--CATGAATAACATTACG 423
DB 537 AACCAAGCCTCAAGGCTTCTGACACTTAAGATTATTCATACCGGAGAGAACTTATGAA 596
QY 424 TGTAAACATGTATTAAGTTTTCATCTGACCTTCAAAATTTAGACAACATTAATTAAA 483
DB 597 TGTGAAGAAATGTGGCAAGCCTTTTAACTCCTCAAACTTACTATATACATAAGATTAAT 656

OY	484	CATCATGGTG	493
Dd	657	CATTCTGGAG	666
RESULT 13			
LOCUS	CN417582	752 bp	mRNA linear EST 16-MAY-2004
DEFINITION	U7000424010604 GRN_EB Homo sapiens cDNA 5', mRNA sequence.		
ACCESSION	CN417582		
VERSION	CN417582.1	GI:47405176	
KEYWORDS	EST.		
SOURCE	Homo sapiens	(human)	
ORGANISM	Homo sapiens		
AUTHORS	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo. 1 (bases 1 to 752) Brandenberger,R., Wei,H., Zhang,S., Lei,S., Murag,J., Fisk,G.J., Li,Y., Xu,C., Fang,R., Guegler,K., Rao,M.S., Mandalam,R., Ilekowski,J and Stanton,L.W. Transcriptome characterization elucidates signaling networks that control human ES cell growth and differentiation Nat. Biotechnol. 22 (6), 707-716 (2004) Contact: Brandenberger R Regenerative Medicine Geron Corporation 230 Constitution Drive, Menlo Park, CA 94025, USA Tel.: 650 473 8658 Fax: 650 473 7760 Email: rbrandenberger@geron.com Insert Length: 752 Std Error: 0.00.		
JOURNAL			
COMMENT	Location/Qualifiers 1..752 /organism="Homo sapiens" /mol_type="mRNA" /db_xref="taxon:9606" /tissue_type="embryonic stem cells, embryoid bodies derived from H1, H7 and H9 cells" /clone_lib="GRN_EB" /note="Oligo dT primed, full-length enriched cDNA library from embryoid body outgrowths derived from hES cell lines H1 (p32), H7 (p29), and H9 (p26) maintained in feeder-free conditions."		
FEATURES			
SOURCE			
ORIGIN			
Query Match	7.4%	Score 92;	DB 7; Length 752;
Beech Local Similarity	52.0%;	Pred. No. 5.7e-08;	
Matches 259;	Conservative 0;	Mismatches 230;	Indels 9; Gaps 2;
OY	7	GAAAGTCGAACAACCAATGCATATCACTTTAATATCTTCTTCATCGACGTCC	66
Dd	72	GAATGTGCCAAGCCTTACCATTCTTCAACACTTACACATAAGTAATTCATCT	131
OY	67	AAAAAGTATATTGGACATATGAAGGGGTGATTAAGCCTTAATCGACATATTATTA	126
Dd	132	GGAGGGAAGCCCTTCAATGTGAAGAATGTGTAAGCTTTAACCAACCTTCAGCCCTT	191
OY	127	GAGCAACATTTAAGAACCCACAGTAATATGACCGCTTAATATGTACAGTGAAGATTGT	186
Dd	192	GCTACACATTAAGTTCATTCATGTTAAAAGAAAACCTTCAAAATGT-----GAAGAATGT	245
OY	187	GATAAAGCATTTTTAGAAAATTCACATTTGGAAAACACATATGTATCATTCATTCGAAAAA	246
Dd	246	GACAAAGCTTTTAACCGATTCTCATCTTACTTAACATCAATGAATTAATTCATTTCTGGAGAG	305
OY	247	AAACCATTCATTTCTGTCAGTGTGTGTGAAGGGGTTAATTCGACAAACACTTAAAAA	306
Dd	306	AAATCTTAACAATGTGAACATATGTGGCAAAGGCTTTAACTGTCTCAACCTTACAAAA	365
OY	307	CATGAATACCCCATCAAAAGTCATTTAATGTACATTTGAAAATTTGTCAAGAAAGCATTT	366
Dd	366	CATGAAGAATTCATATCTGAGAGAGAAACCTTCAAAATGTGAAGAATGTGGCAAAGCCTTT	425

Qy	367	TATAAAGT---	CAATCTTTAAGCATCATATATATATCTGTTCAAGAAAAACATTAAAG	423
Db	426	AATGTGCTTACACCTTACTACATTAAGATGATTCACTAGGAGAAACCTTACAA	485	
Qy	424	TGTAAACATGATTAATAAAGTTTTCACCTGCACCTTCAAAATTAAGCAACAATTAAATTAA	483	
Db	486	TGTGAAGATGTGGCAAAAGCCTTTAAACACCTCTCAAAACCTTACTATTAAGATTAAT	545	
Qy	484	CATCATGTGTGATCTCCT	501	
Db	546	CATACTGGAGAAACCT	563	
RESULT 14				
B99387				
DEFINITION	B99387	593 bp	DNA	linear
LOCUS	CIT-HSP-2281A10.TF	CIT-HSP	Homo sapiens	genomic clone 2281A10,
ACCESSION	B99387			genomic survey sequence.
VERSION	B99387			
KEYWORDS	B99387.1	GI:3027197		
SOURCE	GS5.			
ORGANISM	Homo sapiens (human)			
REFERENCE	Homo sapiens			
AUTHORS	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo. 1 (bases 1 to 593)			
TITLE	Adams,M.D., Rounsley,S.D., Zhao,S., Field,C.E., Bass,S., Linher,K. Golden,K., Berry,K., Ganger,D., Sun,E., Wible,C., Shizuya,H., Simon,M., and Venter,J.C. Use of a random BAC End Sequence Database for Sequence-Ready Map Building (1998)			
JOURNAL	Unpublished (1998)			
COMMENT	Contact: Mark Adams Department of Eukaryotic Genomics The Institute for Genomic Research 9712 Medical Center Dr., Rockville, MD 20850, USA Tel: 301 838 0200 Fax: 301 838 0208 Email: mdadams@tigr.org Clones are available from Research Genetics (info@reagen.com). BAC end search page: http://www.tigr.org/cdb/humgen/bac_end_search/bac_end_search.html Seq primer: M13-21; Class: BAC ends.			
FEATURES				
SOURCE	Location/Qualifiers			
	1..593			
	/organism="Homo sapiens"			
	/mol_type="genomic DNA"			
	/db_xref="GDB:7147044"			
	/db_xref="taxon:9606"			
	/clone="2281A10"			
	/sex="Male"			
	/cell_type="Sperm"			
	/clone_id="CIT-HSP"			
	/note="Vector: pBelobac11; Site_1: HindIII; Site_2: HindIII"			
ORIGIN				
Query Match	7.4%;	Score 91;	DB 8;	Length 593;
Best local Similarity	54.4%;	Pred. No. 8.8e-08;		
Matches 231;	Conservative 0;	Mismatches 185;	Indels 9;	Gaps 2
Qy	69	AAATTAATTGGCACAATATGAAGGTTGGATGAATGAAGCCCTTAATATGACATCATCTTTATG	128	
Db	47	AAAAGAAACCCCTACAAATGTGAAGATGTGGCAAGCTTTTGCCAGCTCTCAACCTTAG	106	
Qy	129	GCAACATTTAAGAACCCACAGTATATGATCGACCGTATTAATATGATCAGTGACGATTTGTA	188	
Db	107	AAAACATGAGATTAATTTACTACTGAGAGAAACCTTACAAATCT-----GAAGATGTGG	160	
Qy	189	TAAAGCATTTTTCAGAAAATCACAATTTGGAAACACATATTTGTATCATTTCGAAAAA	248	

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 8, 2005, 22:51:59 ; Search time 5292.91 Seconds
(without alignments)
11342.728 Million cell updates/sec

Title: US-09-831-804-2
Perfect score: 1239
Sequence: 1 atgcagtgaaagtcagcaaac.....catcagtgattctcgataa 1239

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 4708233 seqs, 24227607955 residues
Total number of hits satisfying chosen parameters: 9416466

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Maximum Match 0%
Listing first 45 summaries

Database :

GenEmbl: *
1: gb_ba: *
2: gb_btg: *
3: gb_in: *
4: gb_om: *
5: gb_ov: *
6: gb_pat: *
7: gb_ph: *
8: gb_pl: *
9: gb_pr: *
10: gb_ro: *
11: gb_sts: *
12: gb_sy: *
13: gb_un: *
14: gb_vi: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1239	100.0	1239	6	BD274350
2	1239	100.0	1239	6	BD274349
3	1235.8	99.7	1239	6	AX489193
4	584.8	47.2	627	6	AR550053
5	438.4	35.4	110000	8	CR382135_06
6	391	31.6	462	6	AR550055
7	228	18.4	110000	8	CR382122_03
8	227.2	18.3	92211	8	CR380953_09
9	217.4	17.5	1560	8	YSCNOFEAT
10	217.4	17.5	1739	8	YSCPTIIIA
11	217.4	17.5	37497	8	YSCPTIIIA
12	199.2	16.1	2133	8	YSCRP026A
13	193	15.6	797	11	CNSO6ING
14	120.6	9.7	110000	8	AE016818_01
15	101	8.2	169063	9	AC138126
16	101	8.2	177299	9	AC073544
17	99.4	8.0	2132	9	AX714042
18	99.4	8.0	2132	9	AK056088
19	99.4	8.0	2196	6	CQ726027

C 20	99.4	8.0	138627	9	AC099500	AC099500 Homo sapi
C 21	99.4	8.0	169500	2	AC138469	AC138469 Homo sapi
C 22	98.2	7.9	2237	9	AK122869	AK122869 Homo sapi
C 23	97.8	7.9	1293	9	D70831	D70831 Homo sapien
C 24	97.8	7.9	235532	9	AC008739	AC008739 Homo sapi
C 25	97.2	7.8	115995	9	AC0011494	AC0011494 Homo sapi
C 26	96.6	7.8	2320	6	AR270491	AR270491 Sequence
C 27	96.6	7.8	2320	6	HSU5376	U5376 Human repre
C 28	96.6	7.8	68304	10	BR000432	BR000432 Mouse DNA
C 29	96.6	7.8	106811	2	CR56705	CR56705 Mus muscu
C 30	95	7.7	981	9	HSZEF	X78932 H. sapiens H
C 31	94.6	7.6	186233	9	AC092329	AC092329 Homo sapi
C 32	94.6	7.6	189317	2	AC024483	AC024483 Homo sapi
C 33	94.2	7.6	292390	2	AC105677	AC105677 Rattus no
C 34	93.6	7.6	1389	9	M27879	M27879 Homo sapien
C 35	93.6	7.6	155439	9	AC008626	AC008626 Homo sapi
C 36	93.4	7.5	199585	10	AC124426	AC124426 Mus muscu
C 37	92.4	7.5	3516	6	CQ726030	CQ726030 Sequence
C 38	92.4	7.5	41153	9	AC016628	AC016628 Homo sapi
C 39	92.4	7.5	158430	2	AC012431	AC012431 Homo sapi
C 40	92	7.4	2873	9	HUMRPLK	M55422 Human Kruep
C 41	92	7.4	2873	11	G28705	G28705 SMS3976 Br
C 42	92	7.4	2873	11	G28705	G18281 human chrom
C 43	92	7.4	156835	9	AC010620	AC010620 Homo sapi
C 44	92	7.4	191082	2	AC141066	AC141066 Homo sapi
C 45	92	7.4	203396	9	AC073210	AC073210 Homo sapi

ALIGNMENTS

RESULT 1
BD274350 1239 bp DNA linear PAT 17-JUL-2003
BD274350
LOCUS
DEFINITION
Candida albicans tflIIA gene (CafIIIA) and the coded CafIIIA
protein.
ACCESSION
BD274350.1 GI:33084118
VERSION
JP 2002531068-A/2.
KEYWORDS
Candida albicans
SOURCE
Candida albicans
ORGANISM
Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
Saccharomycetales; mitosporic Saccharomycetales; Candida.
REFERENCE
Pallier, F.B., Camier, S. and Sentenac, A.
Candida albicans tflIIA gene (CafIIIA) and the coded CafIIIA
protein. Patent: JP 2002531068-A/2 24-SEP-2002;
JOURNAL
AVENTIS PHARMA SA
COMMENT
OS Candida albicans
PN JP 2002531068-A/2
PD 24-SEP-2002
PP 09-NOV-1999 JP 2000581204
PR 10-NOV-1998 FR 98/14147
PI FLORENCE BORDON PALLIER, SYLVIE CAMIER, ANDRE SENTENAC PC
C12N15/09, C12N15/09, A61K39/00, A61K5/00, A61P31/10, C07K14/40, PC
C07K16/14,
PC C12N1/19, C12N1/21, C12P21/02, C12Q1/02, G01N33/15, G01N33/50, PC
G01N33/53,
PC G01N33/566, G01N33/569// (C12N15/09, C12R1.725), C12N15/00, C12N15/00, C12R1.725)
PC 00,
PC (C12N15/00, C12R1.725)
CC Candida albicans tflIIA gene (CafIIIA) and the coded CafIIIA
protein
FH Key
FT CDS Location/Qualifiers
1..1239
Location/Qualifiers
1..(1236).
/organism="Candida albicans"
/mol_type="genomic DNA"
/db_xref="taxon:5476"

FEATURES
source
ORIGIN
Query Match 100.0%; Score 1239; DB 6; Length 1239;

Best Local Similarity 100.0%; Pred. No. 5.7e-190;
Matches 1239; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

QY 1 ATGAGTGAAGTGAACCAATCGATATCATCTTTAATATCTTCTTCTTCAATCA 60
Db 1 ATGAGTGAAGTGAACCAATCGATATCATCTTTAATATCTTCTTCTTCAATCA 60
QY 61 CGTCCCAAAAAGTATTTTGCAATATGAAGGGGTGTATTAAGCCATTAATCGACATCA 120
Db 61 CGTCCCAAAAAGTATTTTGCAATATGAAGGGGTGTATTAAGCCATTAATCGACATCA 120
QY 121 TTATTAGAGCAATTTAAGAACCCACAGTAATGACCCGATTAATAGTAAGTGAAC 180
Db 121 TTATTAGAGCAATTTAAGAACCCACAGTAATGACCCGATTAATAGTAAGTGAAC 180
QY 181 GATTGTGATTAAGCATTTTTCAGAAAATCATTTGGAAAACATATTTGATCATCTCC 240
Db 181 GATTGTGATTAAGCATTTTTCAGAAAATCATTTGGAAAACATATTTGATCATCTCC 240
QY 241 GAAAAAAAACATTCATTTGTCAGTGTGTGAAGGGGTAAATTTCTGCAACAACCTTG 300
Db 241 GAAAAAAAACATTCATTTGTCAGTGTGTGAAGGGGTAAATTTCTGCAACAACCTTG 300
QY 301 AAAAGCATGAATATCCCATCAAGTCAATTTAAATGTATCAATTTGAAAATTTGTCAAGA 360
Db 301 AAAAGCATGAATATCCCATCAAGTCAATTTAAATGTATCAATTTGAAAATTTGTCAAGA 360
QY 361 GCATTTTAAACATCATCTTTAAGACATCATATATTTCTGTCATGAAAACATTA 420
Db 361 GCATTTTAAACATCATCTTTAAGACATCATATATTTCTGTCATGAAAACATTA 420
QY 421 ACGTGAACCATGTATTAAGTTTTCATCGACCTTCAAAATTTAGCAACAATAATTA 480
Db 421 ACGTGAACCATGTATTAAGTTTTCATCGACCTTCAAAATTTAGCAACAATAATTA 480
QY 481 AAAACATCATGTGATCTCTGTTTATCAATGTATCATCTGTTGTTTAAATTTTC 540
Db 481 AAAACATCATGTGATCTCTGTTTATCAATGTATCATCTGTTGTTTAAATTTTC 540
QY 541 CAAACTTGGTCAGTATTTAACAATTTCAATTAACAACATGATCAAAAACCTTAAATGCTC 600
Db 541 CAAACTTGGTCAGTATTTAACAATTTCAATTAACAACATGATCAAAAACCTTAAATGCTC 600
QY 601 AAATGTGTAAAGTTGTGTGGGAAAAAGTTTATCTTCAATATGTTAAGTCATGAT 660
Db 601 AAATGTGTAAAGTTGTGTGGGAAAAAGTTTATCTTCAATATGTTAAGTCATGAT 660
QY 661 GATTCTACCATGATCAAAATATGAGACTTGTGATATGATGTGGGAAATTTGCAAG 720
Db 661 GATTCTACCATGATCAAAATATGAGACTTGTGATATGATGTGGGAAATTTGCAAG 720
QY 721 AAAAATGAATTAAGTGAACATTAATATCTTCCATGATGTAATATCCCTGATGATTA 780
Db 721 AAAAATGAATTAAGTGAACATTAATATCTTCCATGATGTAATATCCCTGATGATTA 780
QY 781 TTAAAGAAACCTGAAGTGAAGAAATTTAGAAACCTATTAAGATCAAGATCGAAATTAAT 840
Db 781 TTAAAGAAACCTGAAGTGAAGAAATTTAGAAACCTATTAAGATCGAAATTAAT 840
QY 841 AATTTCATGAATTAAGAAACAGAAATTTAAAGTGAAGAAAGATTAAGAGTGAAGAA 900
Db 841 AATTTCATGAATTAAGAAACAGAAATTTAAAGTGAAGAAAGATTAAGAGTGAAGAA 900
QY 901 GATAGCTAGATGAAGAAAGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 960
Db 901 GATAGCTAGATGAAGAAAGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 960
QY 961 ATAAATCATTTACTGCTTCTTGAAGGTTCAAGAGTGTCTTAACTTAATTTCTGAAT 1020
Db 961 ATAAATCATTTACTGCTTCTTGAAGGTTCAAGAGTGTCTTAACTTAATTTCTGAAT 1020
QY 1021 AGTGGAGAGATCATTTGCTTAAGAAATATTTGTATGAATGTTTCTAGAGAAAT 1080
Db 1021 AGTGGAGAGATCATTTGCTTAAGAAATATTTGTATGAATGTTTCTAGAGAAAT 1080

```

```

Db 1021 AGTGGAGAGATCATTTGCTTAAGAAATATTTGTATGAATGTTTCTAGAGAAAT 1080
QY 1081 GATTATCGTCGACATTTGAAATGCGATGATGATTAATTTACAAAGATGATCATTTCTTA 1140
Db 1081 GATTATCGTCGACATTTGAAATGCGATGATGATTAATTTACAAAGATGATCATTTCTTA 1140
QY 1141 AATAGTATGAAAAAGAAAGAACTCCAGAGGTGAACCATTTGTTAAAAAGCCAGATG 1200
Db 1141 AATAGTATGAAAAAGAAAGAACTCCAGAGGTGAACCATTTGTTAAAAAGCCAGATG 1200
QY 1201 GATTATTTGCCAAATGAAGAAACATCACTGATGATTTCTCGATTA 1239
Db 1201 GATTATTTGCCAAATGAAGAAACATCACTGATGATTTCTCGATTA 1239

```

```

RESULT 2
BD274349
LOCUS BD274349 2060 bp DNA linear PAT 17-JUL-2003
DEFINITION Candida albicans tfl1111 gene (Catf1111) and the coded CATF1111 protein.
ACCESSION BD274349
VERSION BD274349.1 GI:33084117
KEYWORDS JP 2002531068-A/1..
SOURCE Candida albicans
ORGANISM Candida albicans
Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
Saccharomycetales; Mitosporic Saccharomycetales; Candida.

```

```

REFERENCE
AUTHORS Pallier, F. B., Camier, S. and Sentenac, A.
TITLE Candida albicans tfl1111 gene (Catf1111) and the coded CATF1111
JOURNAL Patent: JP 2002531068-A 1 24-SEP-2002;
AVENTIS PHARMA SA
OS Candida albicans
PN JP 2002531068-A/1
PD 24-SEP-2002
PF 09-NOV-1999 JP 2000581204
PR 10-NOV-1998 FR 98/14147
PI FLORENCE BORDON PALLIER, SYLVIE CAMIER, ANDRE SENTENAC PC
C12N15/09, C12N15/09, A61K45/00, A61P31/10, C07K14/40, PC
C07K16/14,
PC C12N1/19, C12N1/21, C12P21/02, C12P1/02, G01N33/15, G01N33/50, PC
G01N33/53,
PC G01N33/56, G01N33/569// (C12N15/09, C12R1:725), C12N15/00, C12N15/00,
C12N15/00, C12R1:725)
CC Candida albicans tfl1111 gene (Catf1111) and the coded CATF1111
CC protein
FH Key Location/Qualifiers
FT source 1..2060
FT 1..2060
FT Location/Qualifiers

```

FEATURES

```

source 1..2060
location/Qualifiers
1..2060
/organism="Candida albicans"
/mol_type="genomic DNA"
/db_xref="taxon:5476"

```

ORIGIN

Query Match 100.0%; Score 1239; DB 6; Length 2060;
Best Local Similarity 100.0%; Pred. No. 5e-190;
Matches 1239; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

QY 1 ATGAGTGAAGTGAACCAATCGATATCATCTTTAATATCTTCTTCTTCAATCA 60
Db 720 ATGAGTGAAGTGAACCAATCGATATCATCTTTAATATCTTCTTCTTCAATCA 779
QY 61 CGTCCCAAAAAGTATTTTGCAATATGAAGGGGTGTATTAAGCCATTAATCGACATCA 120
Db 780 CGTCCCAAAAAGTATTTTGCAATATGAAGGGGTGTATTAAGCCATTAATCGACATCA 839
QY 121 TTATTAGAGCAATTTAAGAACCCACGATATGATGACCGTATTAATGTACAGTGAC 180
Db 840 TTATTAGAGCAATTTAAGAACCCACGATATGATGACCGTATTAATGTACAGTGAC 899

```

QY 181 GATTGATTAAGCATTTTTCAGAAAAATCATTGGAAAACATAATTGATACATATCC 240
DB 900 GATTGATTAAGCATTTTTCAGAAAAATCATTGGAAAACATAATTGATACATATCC 959
QY 241 GAAAAAAACCATTCATTGTTTCAGTGTGTGTAAAGGGTTAATTCAGCAACCTTG 300
DB 960 GAAAAAAACCATTCATTGTTTCAGTGTGTGTAAAGGGTTAATTCAGCAACCTTG 1019
QY 301 AAAAGCATGAATCAACCCATCAAGAGTCATTTAAATGTCATTGAAAAATTCAGAA 360
DB 1020 AAAAGCATGAATCAACCCATCAAGAGTCATTTAAATGTCATTGAAAAATTCAGAA 1079
QY 361 GCATTTTAAACATCAATCTTTAAAGCATCATATATTTCTGTCATGAAAAAATTC 420
DB 1080 GCATTTTAAACATCAATCTTTAAAGCATCATATATTTCTGTCATGAAAAAATTC 1139
QY 421 ACGTGTAAACATGTATTAAGTTTTCATCGACCTTCAAAATTAAGCAACATTAATTA 480
DB 1140 ACGTGTAAACATGTATTAAGTTTTCATCGACCTTCAAAATTAAGCAACATTAATTA 1199
QY 481 AAACATCATGATGATCTCTGCTTATCATGTGATCATCTGTTGTTTAAATTC 540
DB 1200 AAACATCATGATGATCTCTGCTTATCATGTGATCATCTGTTGTTTAAATTC 1259
QY 541 CAACTTGTGATGATTAATTCATTAATTAAGCATCATATATTTCTGTCATGAAAAA 600
DB 1260 CAACTTGTGATGATTAATTCATTAATTAAGCATCATATATTTCTGTCATGAAAAA 1319
QY 601 AAATGTGTAAAGGTGTGTGTGAAAAAAGTTTATCTTCATATGTTAACTGAT 660
DB 1320 AAATGTGTAAAGGTGTGTGTGAAAAAAGTTTATCTTCATATGTTAACTGAT 1379
QY 661 GATTTCACCATGATCAAAAATATGACCTTGATTTATGTGATGTGGAAAAATTCGAA 720
DB 1380 GATTTCACCATGATCAAAAATATGACCTTGATTTATGTGATGTGGAAAAATTCGAA 1439
QY 721 AAAAAGAAATTAAGTGAACATTAATATCTTCATGATGATGATCTGATGATTTA 780
DB 1440 AAAAAGAAATTAAGTGAACATTAATATCTTCATGATGATGATCTGATGATTTA 1499
QY 781 TTTAAAGAAATGAGTGAAGAAAAATTAGAACCTTATGATCAAGATCGAAATTAAT 840
DB 1500 TTTAAAGAAATGAGTGAAGAAAAATTAGAACCTTATGATCAAGATCGAAATTAAT 1559
QY 841 AAATTCATGAATTTAGAAACAGAGAAATTAAGTGAAGAGATGAAGAAATGAAGA 900
DB 1560 AAATTCATGAATTTAGAAACAGAGAAATTAAGTGAAGAGATGAAGAAATGAAGA 1619
QY 901 GATAGCTAGATGAAAAAAGAGATGATGATGATGATGATGATGATGATGATGATGAT 960
DB 1620 GATAGCTAGATGAAAAAAGAGATGATGATGATGATGATGATGATGATGATGATGAT 1679
QY 961 ATAAATCATTTAATCTGCTTCTTGAAGTTCAGAAAGTTCCTAACTTAATTCGAT 1020
DB 1680 ATAAATCATTTAATCTGCTTCTTGAAGTTCAGAAAGTTCCTAACTTAATTCGAT 1739
QY 1021 AGTGGGAAGAGATCAATGCTCTTGAAGATTAATGATGATGATGATGATGATGAT 1080
DB 1740 AGTGGGAAGAGATCAATGCTCTTGAAGATTAATGATGATGATGATGATGATGAT 1799
QY 1081 GATTTACGTGACATTTGAAGTGAAGATGATTAATTAAGAAATTCGATCTTA 1140
DB 1800 GATTTACGTGACATTTGAAGTGAAGATGATTAATTAAGAAATTCGATCTTA 1859
QY 1141 AATAGTATAGAAAAAGAGAAATCTCAGAAAGTGAACATTTGTTTAAAGCAGAGAT 1200
DB 1860 AATAGTATAGAAAAAGAGAAATCTCAGAAAGTGAACATTTGTTTAAAGCAGAGAT 1319
QY 1201 GATTATTTGCCAAATGAACATCATGATATTTCTGATTA 1239
DB 1920 GATTATTTGCCAAATGAACATCATGATATTTCTGATTA 1958

RESULT 3
AX489193 1239 bp DNA linear PAT 16-AUG-2002
LOCUS AX489193
DEFINITION Sequence 6493 from Patent WO02053728.
ACCESSION AX489193
VERSION AX489193.1 GI:22323205
KEYWORDS
SOURCE
ORGANISM
Candida albicans
Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
Saccharomycetales; Mitoportic Saccharomycetales; Candida.
REFERENCE
1 Roemer, T., Jiang, B., Boone, C., Bussey, H. and Olsen, K. L.
AUTHORS Gene disruption methodologies for drug target discovery
TITLE Patent: WO 02053728-A 6493 11-JUL-2002;
JOURNAL Elitza Pharmaceuticals, Inc. (US)
FEATURES
source
1..1239
Location/Qualifiers
/organism="Candida albicans"
/mol_type="unassigned DNA"
/db_xref="taxon:5476"
ORIGIN
Query Match 99.7%; Score 1235.8; DB 6; Length 1239;
Best Local Similarity 99.8%; Pred. No. 1.9e-189;
Matches 1237; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 1 ATGAGTGAAGATGACGAAACCAATGATATCATCTTTAATATCTTCTTCATCA 60
DB 1 ATGAGTGAAGATGACGAAACCAATGATATCATCTTTAATATCTTCTTCATCA 60
QY 61 CGTCCAAAAGATATTTTGACATATGAAAGGGTGTATTAAGCCTATATGACCATCA 120
DB 61 CGTCCAAAAGATATTTTGACATATGAAAGGGTGTATTAAGCCTATATGACCATCA 120
QY 121 TTTATGAGCAACATTTTAAAGAACCCACAGTATGATCGACCGTATTAATGATACGTGAC 180
DB 121 TTTATGAGCAACATTTTAAAGAACCCACAGTATGATCGACCGTATTAATGATACGTGAC 180
QY 181 GATTGATTAAGCATTTTTCAGAAAAATCATTG3AAAACATATGATATCATATCC 240
DB 181 GATTGATTAAGCATTTTTCAGAAAAATCATTG3AAAACATATGATATCATATCC 240
QY 241 GAAAAAAACCATTCATTGTTTCAGTGTGTGTAAAGGGTTAATTCAGCAACCTTG 300
DB 241 GAAAAAAACCATTCATTGTTTCAGTGTGTGTAAAGGGTTAATTCAGCAACCTTG 300
QY 301 AAAAGCATGAATCAACCCATCAAGAGTCATTTAAATGTCATTGAAAAATTCAGAA 360
DB 301 AAAAGCATGAATCAACCCATCAAGAGTCATTTAAATGTCATTGAAAAATTCAGAA 360
QY 361 GCATTTTAAACATCAATCTTTAAAGCATCATATATTTCTGTCATGAAAAAATTC 420
DB 361 GCATTTTAAACATCAATCTTTAAAGCATCATATATTTCTGTCATGAAAAAATTC 420
QY 421 ACGTGTAAACATGTATTAAGTTTTCATCGACCTTCAAAATTAAGCAACATTAATTA 480
DB 421 ACGTGTAAACATGTATTAAGTTTTCATCGACCTTCAAAATTAAGCAACATTAATTA 480
QY 481 AAACATCATGATGATCTCTGCTTATCATGTGATCATCTGTTGTTTAAATTC 540
DB 481 AAACATCATGATGATCTCTGCTTATCATGTGATCATCTGTTGTTTAAATTC 540
QY 541 CAACTTGTGATGATTAATTCATTAATTAAGCATCATATATTTCTGATCAAAATTCCT 600
DB 541 CAACTTGTGATGATTAATTCATTAATTAAGCATCATATATTTCTGATCAAAATTCCT 600
QY 601 AAATGTGTAAAGGTGTGTGTGAAAAAAGTTTATCTTCATATGTTAACTGAT 660
DB 601 AAATGTGTAAAGGTGTGTGTGAAAAAAGTTTATCTTCATATGTTAACTGAT 660
QY 661 GATTTCACCATGATCAAAAATATGACCTTGATTTATGTGATGTGGGAAATTCGAA 720

Db 661 GATTCACCATGATCAAAATATGACTTGTGATTATTTGANTGCGGAAATTTGCAAG 720
QY 721 AAAAATGAATTAGTGAACATTAATAATCTTCATGATGATTAATCCCTGATGATTA 780
Db 721 AAAAATGAATTAGTGAACATTAATAATCTTCATGATGATTAATCCCTGATGATTA 780
QY 781 TTTAAAGAACTGAAGTGAAGAAATTTAGAACCTTAATGATCAAGATCGAAATTTAAAT 840
Db 781 TTTAAAGAACTGAAGTGAAGAAATTTAGAACCTTAATGATCAAGATCGAAATTTAAAT 840
QY 841 AATTGCAATGAATTAAGAACAGAAATTTAAAGTGAAGAAAGATTAAGAAAGATGAAGAA 900
Db 841 AATTGCAATGAATTAAGAACAGAAATTTAAAGTGAAGAAAGATTAAGAAAGATGAAGAA 900
QY 901 GATAGCTAGATGAAAAAGAAAGATGATTTAGATGACACTCAATGTCAGCTCAAGATCA 960
Db 901 GATAGCTAGATGAAAAAGAAAGATGATTTAGATGACACTCAATGTCAGCTCAAGATCA 960
QY 961 ATTAATTCATTATCTGCTCTTTGGAAGTTCAAAGAGTTTCTAACTTATTTGAAAT 1020
Db 961 ATTAATTCATTATCTGCTCTTTGGAAGTTCAAAGAGTTTCTAACTTATTTGAAAT 1020
QY 1021 AGTGGAGAGATCAATGCTCTTAAGAAATTAATTTGATGAATGTTTTCTAGAGAAATAT 1080
Db 1021 AGTGGAGAGATCAATGCTCTTAAGAAATTAATTTGATGAATGTTTTCTAGAGAAATAT 1080
QY 1081 GATTTACGTGACATTTGAAATGTCATGATTAATTTTCAAGAAATTTGATGATCTTA 1140
Db 1081 GATTTACGTGACATTTGAAATGTCATGATTAATTTTCAAGAAATTTGATGATCTTA 1140
QY 1141 AATAGTATGAAAAAGAAAGAACTCCAGAAAGTGAACATTTGTTAAAAAGCCAGAGATG 1200
Db 1141 AATAGTATGAAAAAGAAAGAACTCCAGAAAGTGAACATTTGTTAAAAAGCCAGAGATG 1200
QY 1201 GATTTATGCGCAATGAATCATGATGATTTCTCGATTA 1239
Db 1201 GATTTATGCGCAATGAATCATGATGATTTCTCGATTA 1239

RESULT 4
ARS50053
LOCUS ARS50053 627 bp DNA linear PAT 08-OCT-2004
DEFINITION Sequence 5184 from patent US 6747137.
ACCESSION ARS50053
VERSION ARS50053.1 GI:53943228
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 627)
AUTHORS Weinstein, K.G. and Bush, D.
TITLE Nucleic acid sequences relating to Candida albicans for diagnostics
JOURNAL Patent: US 6747137-A 5184 08-JUN-2004;
FEATURES
source
1..627
/organism="Unknown"
/mol_type="genomic DNA"

ORIGIN
Query Match 47.2%; Score 584.8; DB 6; Length 627;
Best Local Similarity 97.4%; Pred. No. 1.3e-84;
Matches 608; Conservative 0; Mismatches 7; Indels 9; Gaps 1;
QY 616 TGTGTTGGAAAAAGGTTTATCTTCAATATGTTAAGTCATGATGATTTCTACATGATC 675
Db 13 TGTGTTGGAAAAAGGTTTATCTTCAATATGTTAAGTCATGATGATTTTACCAATGATC 72
QY 676 AAAATATGACTTGTGATTTGATGTTGGGAAATTTGCAAGAAAAATGATTAATGTT 735
Db 73 AAAATATGACTTGTGATTTGATGTTGGGAAATTTGCAAGAAAAATGATTAATGTT 132
QY 736 GAACATTAATATATCTTCATGATGATTAATCCCTGATGATTTATTAAGAAACTGAA 795

Db 133 GAACATTAATTAATCTTCATGATGATTAATCCCTGATGATTTATTAAGAAACTGAA 192
QY 796 GTGAAGAAATTTAGAACCTTATAGATCAAGATCGAAATTTAAATTTGATGAATTA 855
Db 193 GTGAAGAAATTTAGAACCTTATAGATCAAGATCGAAATTTAAATTTGATGAATTA 252
QY 856 GAACAGAGAAATTTAAAGTGGAGAGATGAAGAAATGAAGATAGTCTAGATGAA 915
Db 253 GAACAGAGAAATTTAAAGTGGAGAGATGAAGAAATGAAGATAGTCTAGATGAA 303
QY 916 AAAAGATGATGTTAGATCAGATCTCAATGTCAGCTCAAGATCAATTAATCAATTTACT 975
Db 304 AGAAGAGATGATTTAGATCAGATCTCAATGTCAGCTCAAGATCAATTAATCAATTTACT 363
QY 976 GCTTCTTTGGAAGTTTCAAGAGTGTCTTAACCTTAATTTGATTAAGTGGAGAAATATC 1035
Db 364 GCTTCTTTGGAAGTTTCAAGAGTGTCTTAACCTTAATTTGATTAAGTGGAGAAATATC 423
QY 1036 AATTGCTTAAGAAATTAATTTGATGAATGTTTTCTAGAGAAATTAATTTAGCTGACAT 1095
Db 424 AATTGCTTAAGAAATTAATTTGATGAATGTTTTCTAGAGAAATTAATTTAGCTGACAT 483
QY 1096 TTTGAATGCGATGATGATTAATTTCAAGAAATTTGATGATCTTTTAAATGATTAAGAAAA 1155
Db 484 TTTGAATGCGATGATGATTAATTTCAAGAAATTTGATGATCTTTTAAATGATTAAGAAAA 543
QY 1156 GAAGAAATCTCCGAAGGTGAACCATTTGTTAAAAAGCCAGAGATTTATTTGCAAT 1215
Db 544 GAAGAAATCTCCGAAGGTGAACCATTTGTTAAAAAGCCAGAGATTTATTTGCAAT 603
QY 1216 GAACATCATGATGATTTCTCGATTA 1239
Db 604 GAACATCATGATGATTTCTCGATTA 627

RESULT 5
CR382135 06/c
WPCOMMENT
Sequence split into 16 fragments LOCUS CR382135 Accession CR382135
Fragment Name Begin End
CR382135_00 1 110000
CR382135_01 100001 210000
CR382135_02 200001 310000
CR382135_03 300001 410000
CR382135_04 400001 510000
CR382135_05 500001 610000
CR382135_06 600001 710000
CR382135_07 700001 810000
CR382135_08 800001 910000
CR382135_09 900001 1010000
CR382135_10 1000001 1110000
CR382135_11 1100001 1210000
CR382135_12 1200001 1310000
CR382135_13 1300001 1410000
CR382135_14 1400001 1510000
CR382135_15 1500001 1592360
Continuation (7 of 16) of CR382135 from base 600001 (CR382135 Debaryomyces hanseani chro
Query Match 35.4%; Score 438.4; DB 8; Length 110000;
Best Local Similarity 63.4%; Pred. No. 1.4e-61;
Matches 724; Conservative 0; Mismatches 406; Indels 12; Gaps 3;
QY 33 ATCTTAATATCTTCTTCTTCAATGATGATTTTCAAGAAATTAATTTGATGAATGAG 92
Db 20471 ATCTTAATATCTTCTTCTTCAATGATGATTTTCAAGAAATTAATTTGATGAATGAG 20412
QY 93 GTGTATTAAGGCTTAATGACATCAATTTTAAAGAACTTTAAAGAACTTAAAGCACTAA 152
Db 20411 CTGCGGAGAGGCTTAATTTTAAAGCAATCGCTTTAGAAACAACTTAAGATCAATAC 20352
QY 153 TGATGACCGTAATTAATGATGATGATGATTTTCAAGAAATCA 212

Db 20351 TGAAGACCTTTAAATGCTCGAAGCTGATCGCAAGATCTCTTAAGAAAGTCACA 20292
QY 213 TTTGGAACACATATGTATC-----ACATTCGGAAAAAAACATTCATGTTCAGT 266
Db 20291 TTTACAGACATTTCTTCTCATGAAGATCAAGAAAGTAAACCTTCCAGTGTCAAC 20232
QY 267 GTGTGTTAAAGGGTTAAATCTCGACACACTGAAAGACATGAATCCCATCAAA 326
Db 20231 TTGGGCAAAAGGTGTAATCCTTACACATTTAAAGACAGAAATTAACATACAA 20172
QY 327 GTCAATTAATGTACATTTGAAAAATGTCAAGAGCATTTTATTAACATCAATCTTAA 386
Db 20171 ATCAATTCATTTGACATTTGAGGGCTGTATGATCATTTTATTAACATCAAGTCA 20112
QY 387 ACATCATATATTTCTGTTTCATGAAAAACATTAAGGTGAACATGTATTAAGTTT 446
Db 20111 ACACCATAGCTATCGGTTTCATGAAAAAAATTTATGCAATAGGTATTAATCGTT 20052
QY 447 CACTGCACCTTCAAAATTTAGCAACAATTAATTAACATCATGTGATCTCTGCTTA 506
Db 20051 TTAACAGACCTATCGATTTGGACACAGCAATTAATTAATCAAGTACTCGCGCTTA 19992
QY 507 TCAATGTATCATCTGTTGTTTAAATTTCAAACTTGGTCAATATTAACATTTCA 566
Db 19991 TCAATGTATCATCAAGGCTGTTTGAATTTTATGAGGTGCTGCTTACATTTGCA 19932
QY 567 TATTAACAACATGTCATCCAAACTTAATGTCTTAATGTGTAAAGTTGTGTGGAA 626
Db 19931 TATTAACAACATGACACCCAAAGATTAAGTCTTAATGTGGAAAGTTGCTGTGATA 19872
QY 627 AAAAGTTATCTTCATATGTTAAGTCATGATGATTCACATGATCAAAATATGAGC 686
Db 19871 GAAAGGCTCTGCTCACATGATATTTCAAGTGAAGAAAGATAGTCAATTTAGAA 19812
QY 687 TTGTGATTAATGTGATGTGGGAAATTTGCAAGAAAAATGATTAAGTGAACATTATA 746
Db 19811 TTGTATTAATGTATCACTTGGAAAAATTTGAAAAAAGCTGATTTGATGATCATTTCA 19752
QY 747 TATCTTCATGATGTATATTCCTGTATGATTTATTAAGGAACTGAGTGAATAAT 806
Db 19751 CATCTTACACGACGAATCTTCTGTATGATTTGAACCCATGCAAGAGACAGTT 19692
QY 807 AGAAGACCTATTA---GATCAAGATCGAAATTAATTAATTTGCATGAATTAAGAACAG 863
Db 19691 AGATTAACCTTAAGGAGAGACTCTACCAATTTGATTTAATCTTTAGAGGATTT 19632
QY 864 GAAATTTAAAGTGAAGAGATGAAGAGATGAAGAGATGCTAG---ATGAAAAAG 920
Db 19631 ACAATCAAAAGGTTTGTGAAGTACCGTGGAGTGAAGAGAGAGAGATGATTAAT 19572
QY 921 AAGTGAATGTATGATCAACTCAATGTCAAGTCAAGATCAATTAATCATTTAAGTCTTC 980
Db 19571 GTCCATATGTTTCAAGCTTCAATTAATCTTTGAATTCGTTGAATTTAATCTTTGAATTC 19512
QY 981 TTTGGAAGGTTCAAGAGTCTTCTAACTTATTTGAAATGAGGAGAGAGATCAATTC 1040
Db 19511 TGGAAAGACATCTATGTATGATCTTATTTGAAATTAATTAATCTGAAAAAGAGTTCCATG 19452
QY 1041 TCTTAAGATTAATTTGATGATGATGTTTCTAGAGATATGATTTAGTGAACATTTGAA 1100
Db 19451 TCCGAAAAAGATTTGATGATGATGTTTCAAGCGTATGATCTTGAAGAGACATTTTAA 19392
QY 1101 ATGGCATGATGATTAATTTACAAAGATGATGATCTTAAATAGTATAGAAAAAGAGA 1160
Db 19391 GTGGCATGATCTCACTTAATAAAGATTAAGAGGTTTAAAAAGTCTTGAAGTGGAGA 19332
QY 1161 AA 1162
Db 19331 AA 19330

RESULT 6
AR550055

LOCUS AR550055 462 bp DNA linear PAT 08-OCT-2004
DEFINITION Sequence 5186 from patent US 6747137.
ACCESSION AR550055
VERSION AR550055.1 GI:53943230
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 462)
AUTHORS Weinstock, K.G. and Bush, D.
TITLE Nucleic acid sequences relating to Candida albicans for diagnostics and therapeutics
JOURNAL Patent: US 6747137-A 5186 08-JUN-2004;
FEATURES
source location/Qualifiers
1..462
/organism="unknown"
/mol_type="genomic DNA"

ORIGIN

Query Match 31.6%; Score 391; DB 6; Length 462;
Best Local Similarity 98.5%; Pred. No. 2.1e-53;
Matches 405; Conservative 0; Mismatches 5; Indels 1; Gaps 1;

QY 1 ATGAGTGAAGTGAACGAAACCAATGATATCATCTTAATATCTTCTTTCATCA 60
Db 52 ATGAGTGAAGTGAACGAAACCAATGATATCATCTTAATATCTTCTTTCATCA 111
QY 61 CGTCCCAAAAGTATTTTGCACATATGAAAGGGTGTATTAAGCTTATTAATGACATCA 120
Db 112 CGTCCCAAAAGTATTTTGTACATATGAAAGGGTGTATTAAGCTTATTAATGACATCA 171
QY 121 TTATTTAGCAACATTTAAGACCCACAGTAATGATCGACCGTAAATGATAGTGCAG 180
Db 172 TTATTTAGCAACATTTAAGACCCACAGTAATGATCGACCGTAAATGATAGTGCAG 231
QY 181 GATTGTATTAAGATTTTTCAGAAATGATCAATTTGAAACATATTTGATCAATTC 240
Db 232 GATTGTATTAAGATTTTTCAGAAATGATCAATTTGAAACATATTTGATCAATTC 291
QY 241 GAAAAAACAATTCATTTGATGATGTGTGTAAGGGGTAAATCTGCAACAACCTTG 300
Db 292 GAAAAAACAATTCATTTGATGATGTGTGTAAGGGGTAAATCTGCAACAACCTTG 351
QY 301 AAAAGCATGAATTCACCATACAAAGTCAATTTAATGTATCAATTTGATCAAGAA 360
Db 352 AAAAGCATGAATTCACCATACAAAGTCAATTTAATGTATCAATTTGATCAAGAA 411
QY 361 GCATTTAT-AAACATCAATCTTAAGACATCATATTAATCTGTCATGA 410
Db 412 ACATTTATTAACATCAATCTTTAAGACATCATATTAATCTGTCATGA 462

RESULT 7
CR382122 03/cWPCOMMENT
Sequence split into 14 fragments LOCUS CR382122 Accession CR382122

Fragment Name	Begin	End
CR382122_00	1	110000
CR382122_01	100001	210000
CR382122_02	200001	310000
CR382122_03	300001	410000
CR382122_04	400001	510000
CR382122_05	500001	610000
CR382122_06	600001	710000
CR382122_07	700001	810000
CR382122_08	800001	910000
CR382122_09	900001	1010000
CR382122_10	1000001	1110000
CR382122_11	1100001	1210000
CR382122_12	1200001	1310000
CR382122_13	1300001	1320834

Continuation 74 of 14) of CR382122 from base 300001 (CR382122 Kluyveromyces fragilis str)

Query Match 18.4%; Score 228; DB 8; Length 110000;
 Best Local Similarity 58.1%; Pred. No. 9.8e-28;
 Matches 443; Conservative 0; Mismatches 310; Indels 9; Gaps 2;

41 TATCTCTCTCTCTCATCAGTCCCAAAAAGATATTTGACATATGAAAGGTGTGATA 100
 15150 TGTCAATTCATCTGTCACAGCGTTCTTAAAGTATATTTGCGATTAATCAAGGTGTGATA 15091

QY 101 AAGCCTAATATGACAGCCATATTATAGACCAATTTAAGAAACCAAGTAAATGATCGAC 160
 DB 15090 AATCGTTCACAAGACGAGTTGTCTCAGGAGATCAGCTCAC--CGTGCATCAGGA 15034

QY 161 CGTATTAATCTACAGTGCAGCATTTGTGATTAAGCATTTTTCAGAAATCACTTTGGAAA 220
 DB 15033 TCMAACCGTTTAAATCGCACCTTGTGGCAAAAGAAATTTGCAAAAATGCAATCTGAATA 14974

QY 221 CACATTTGTATCACTTCGCAAAAACATTCATGTTTCAAGTGTGTGTAAGGGG 280
 DB 14973 GACACATGTTCTCTCTCATCAGATGATTAACCTTTCACATGTTCTATATGTGGAAAGGTG 14914

QY 281 TTAAATCTGACACACATTTGAAAAGACATGAATCCCATCAAAAGTCAATTTAAATGTA 340
 DB 14913 TCACAACGACACAGATTAAAGAGACATGAATTTACATACGAAAGTCTTCCACTGT 14854

QY 341 CATTTGAAAATTTGTCAAGAGCATTTTATTAACATCAATCTTTAGACATCATATATTAT 400
 DB 14853 CATATGAGGAGATGTAAAGCATATCTTACAAAATCCACACTCCCATCGCATATCTGT 14794

QY 401 CGTTTCATGAAAAACATTAAGGTGTAACATGTAATTAAGTTTTCATCGACCTTCA 460
 DB 14793 CGTTTCAGAAAAGAACTTACCTGCTCCCTCCTGTAACAGAGTTCCAAAAGGCTTATC 14734

QY 461 AATTAGACACATTAATTTAAACATCATGTGATCTCTGC-----TTATCATATGTG 514
 DB 14733 GATTGAAGATCACATAGATTAAGCATCAACCCGAAATCTACCGAATGTACATGTG 14674

QY 515 ATCATCTGTTGTTTAAAAATTTCCAACTTGTGATGATTTACATTTCAATTAATTAAC 574
 DB 14673 ACTTCTCTTCATGTAACCGACGATATTTCCACTGTGCTTCTTACAAACATATTTAAAC 14614

QY 575 AACTGATCCAAAACCTTAATGTCCTAATGTGTAAAGTGTGTGGAAGAAAAGTT 634
 DB 14613 AATGCCATCGAAATTAACCATGCTCATTTGTGGGAAACCTTGTGTGTAATCGGTT 14554

QY 635 TATCTTCATATGTTAATGATCATGATTTCTACATGATCAAAATATGAGCTTGATTT 694
 DB 14553 TACGGAATCATATGATGATTTCTATGACGAGATTTGGTCTACTAAGATTTGGAATGTTCTT 14494

QY 695 ATTTGATGTGGGAAATTTGCAAGAAATGAATTAAGTAACTTAATTAATCTTCC 754
 DB 14493 CATGTCAGATACATCATTTGGAAGAAATCTCAGCTAGTACACATTTAAGAAAGTTC 14434

QY 755 ATGATGTATATCCCTGATGATTTATTAAGAAATCTGAAG 796
 DB 14433 ACMAAGATGTAAATCCAGAAATGCTATTTACTGCCGACGAGC 14392

RESULT 8
 CR380953 09/c
 WPCOMMENT
 Sequence split into 10 fragments LOCUS CR380953 Accession CR380953

Fragment Name	Begin	End
CR380953_00	1	110000
CR380953_01	100001	210000
CR380953_02	200001	310000
CR380953_03	300001	410000
CR380953_04	400001	510000
CR380953_05	500001	610000
CR380953_06	600001	710000
CR380953_07	700001	810000
CR380953_08	800001	910000
CR380953_09	900001	992211

Continuation (10 of 10) of CR380953 from base 900001 (CR380953 Candida glabrata strain C

Query Match 18.3%; Score 227.2; DB 8; Length 92211;
 Best Local Similarity 57.8%; Pred. No. 1.4e-27;
 Matches 447; Conservative 0; Mismatches 318; Indels 9; Gaps 2;

44 CTCTCTCTCTCTCATCAGTCCCAAAAAGTATTTGACATATGAAAGGTGTGATAAG 103
 DB 60262 CATCTGATTCAGATGTAAGGTGTGAAGACATCTTTTGAATATGATATATGTCAGAG 60203

QY 104 CCTATATATGACCATCATTTATAGCAACATTTAAGAAACCAAGTAAATGATCGACGT 163
 DB 60202 CCTTTTCAAGGCCGCTCTGCTCAGAGAACACCAAGATATC--AGTGCATTTGGAGAGA 60146

QY 164 ATTAATGTACAGTGCAGCATTTGTGATTAAGCATTTTTCAGAAATCAATTTGGAACAC 223
 DB 60145 AACCTTGAAGATGTAAATCAATGTGAAGATTATTAATCAAAAAGATTCACCTTAGAGAGAC 60086

QY 224 ATATTGTATCATCTTCGAAAAAAACCATTTCCATTTGTTCAGTGTGTGTAAGGGCTTA 283
 DB 60085 ACTTATACACACACAGATGAAAGCCGTTTATTTGTTCAATTTTGTGGAAAGGGCTTA 60026

QY 284 ATTCGACACACTTGAAGAACATGAATCACCCATACAAAGTCAATTTAAATGTATCAT 343
 DB 60025 TTACTAGGCAACATGTAAGAGCATGAGTACTCACACCAATCTTCAATTTGTAT 59966

QY 344 TTGAAAATTTGTCAAGACATTTTATTAACATCAATCTTTAAGACATCATATATTATCTG 403
 DB 59965 ATGAAGTTGTATGAGAGTTTCTACAGACCTTCATTTAAGGGCTCATATTTTAGCAG 59906

QY 404 TTTCAGAAAAACATTAACGTGTAAACAATGTAATTAAGTTTTCAGTGTGTGTAAGGCTTCAAAAT 463
 DB 59905 TTCAATTTCAAAAGTCTAAATGTCAAGATGCACMAAAGTTTTCMAAGACCTTACAGAC 59846

QY 464 TAGCAACATTAATTAATTAACATCAT-----GGTGAATCTCCCTTATCATGATGATC 517
 DB 59845 TCMAAAATCATATGACCAACCAACATTAATCCAGATGTTGTCAATGCTTAATATATCT 59786

QY 518 ATCTGTTGTTTAAAAATTTCCAACTTGTGATGATTTCAATTTCAATTAATTAACAC 577
 DB 59785 TCAGTGTGTTGTTCTAAGAGTTTCAAAACATGATGCTGCGTTAAGATTAATTAATTAAG 59726

QY 578 TGCATCCAAAACCTTAATGTCCTTAATGTGTGAAGTTGTGTGGAAAAAGTTTAT 637
 DB 59725 ATCATCCGAAGTTAAAGCCCTATATGAGTAAACCTTGTGTGGGAAAGCGTTTAA 59666

QY 638 CTTCATATGTTAAGTATGATGATTTCTACATGATCAAAATATGAGCTTGATTT 697
 DB 59665 ATATGACATGAAATTCACAGATGAACCTTGTTCAGAAATGTGAAGTGTCTATAT 59606

QY 698 GTGATGTGGGAAATTTGCAAGAAATGAATTAAGTGAACATTTATATATCTTCATG 757
 DB 59605 GTTAATGATCAGTCAATTTCTAAGAAATTAAGAACTTGTGATCATTAACGACCACTCATTA 59546

QY 758 ATGTAATATCCCTGATGATTTATTAAGAAACTGAAGTGAAGAAATTAAGAGA 811
 DB 59545 GTGAAGAAATCCGCGCTATCTATTAGAAAGAGTGTGCTTATGTGAAA 59492

RESULT 9
 YSCNOFEAT
 LOCUS YSCNOFEAT 1560 bp DNA linear PLN 27-APR-1993
 DEFINITION Saccharomyces cerevisiae transcription factor IIIA and RNA polymerase subunit RPB6 genes.
 M90638
 ACCESSION M90638.1 GI:1172046
 VERSION RNA polymerase; transcription factor IIIA.
 KEYWORDS Saccharomyces cerevisiae (baker's yeast)
 SOURCE Saccharomyces cerevisiae
 ORGANISM Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes; Saccharomycetaceae; Saccharomycetaceae; Saccharomyces.
 REFERENCE 1 (bases 1 to 1560)
 Mochlyk, N.A. and Young, R.A.
 Genes encoding transcription factor IIIA and the RNA polymerase

common subunit RPB6 are divergently transcribed in *Saccharomyces cerevisiae*
Proc. Natl. Acad. Sci. U.S.A. 89 (9), 3999-4003 (1992)

JOURNAL
MEDLINE
PUBMED

92237295
1570325

COMMENT Original source text: *Saccharomyces cerevisiae* (strain S288C) DNA.
FEATURES
source
1. .1560
Location/Qualifiers
/organism="Saccharomyces cerevisiae"
/mol_type="genomic DNA"
/strain="S288C"
/db_xref="taxon:4932"

ORIGIN

chromosome 14.

Query Match 17.5%; Score 217.4; DB 8; Length 1560;

Best Local Similarity 59.4%; Pred. No. 1.4e-25;

Matches 409; Conservative 0; Mismatches 271; Indels 9; Gaps 2;

62 GTCCAAAAAGTATTTTGCATATGAAAGGTTGATTAAGCCTATTAATCGACATCAT 121

368 GGCAGAAAGCATATTTCTGTGATATGATGGCTGTGATTAAGCATTTACAGACCTTCAA 427

122 TATTAGAGCAACATTTAAGAACCCAGTAATATGACCGTATTAATGTACAGTGAGC 181

428 TTTTGACTGAACCAATTAAGCGTACATCAGGGTTTA--AGAGCATTTCACTGTGATA 484

182 ATTGTGATTAAGCATTTTTCAGAAATTCATATTGAAACATATTTGATCATCTCCG 241

485 AGGTGCAAAATCTCTGTATTAAGAGTACTTAAAGAACACTGTATACGATCTCG 544

242 AAAAAAACCATTCATTTGTCAGTGTGTGTAAGGGTTAATTCGCAACACTTGA 301

545 ATACGAAACATTCATATGTTCTTATTTGTGAAAGAGAGAGACCTCGCAGCACTGA 604

302 AAAGACATGAATCAATCCATATCAAGTCAATTTAATGATCAATTTGAAATTTGCAAG 361

605 AGGACACAGAAAGTACGATACCAAAATCTTCAATTTGTCCAGAGAAAGATGCAACTCC 664

362 CATTTATTAACATCATCTTTAAGACATATATATTTATCTGTTCAATGAAACATTTA 421

665 GATTTCTAAGCAATCCCAATTAAGGACATATTTATCTGTTCAATTAATCAATTA 724

422 CGTGAACATATGTAATTAAGTTTCACTGACCTTCAAAATTAAGCAACATTAATTA 481

725 CCGTCCACACTGCAATTAAGCTTTCAAGGCGCATATAGGCTAAGAAATCACTTTCTA 784

482 AACATCA-----TGGTGAATCTCTGCTTATCAATGTGATCAATCTGTTGTTTAA 535

785 AACATCAAGATCTGAGGTGAAATATCTTAACCAATGTAATTTTGTGTTGTTGAAAG 844

536 ATTTCCAACTTGGTCAATTTTCAATTTAATTAACCACTGCATCCAAAATTTAAT 595

845 AGTTTGGATATGATGTCATATTCATTCGATTAATAAAATGATCAATCTTAATTTAAAT 904

596 GTCCATAATGTGTAAGGTTGTTGGGAAAAAAGTTTATCTTCAATATATTTAAGTC 655

905 GTCTTATTTTGAACCAACATGTGTGGGAAAAATGTTTCAAAATGACATATTAATTC 964

656 ATGATGATTTACATGATCAAAATATGACTTGTGATTAATTTGATGTGGGAAAAATTTG 715

965 ATAGCGATCACTAGTAACAAAATTTGAAAGTGTATATATGTCCGATATATGCTTTT 1024

716 CAAAGAAAAATGAATTAAGTTGAACATTAAT 744

1025 CTAGAAAAATGATCTTCTCAGCATTAAT 1053

RESULT 10
YSCFFIIIA

YSCFFIIIA 1739 bp DNA linear PLN 13-SEP-1996

LOCUS *Saccharomyces cerevisiae* transcription factor IIAA (TFIIIA) gene,

DEFINITION complete cds.

ACCESSION

M80611

VERSION M80611.1 GI:172902
KEYWORDS transcription factor IIAA; zinc-finger protein; zinc-finger transcription factor.

SOURCE *Saccharomyces cerevisiae* (baker's yeast)

ORGANISM *Saccharomyces cerevisiae*

Eukaryota; Fungi; Ascomycota; *Saccharomycotina*; *Saccharomycetes*;

Saccharomycetales; *Saccharomycetaceae*; *Saccharomycetes*.

REFERENCE 1 (bases 1 to 1739)

Archambault, J., Milne, C.A., Schapert, K.T., Baum, B., Friese, J.D.

and Segall, J.

The deduced sequence of the transcription factor TFIIIA from

Saccharomyces cerevisiae reveals extensive divergence from *Xenopus*

TFIIIA

JOURNAL

MEDLINE

PUBMED

92147684

1737784

FEATURES

source

Location/Qualifiers

1. .1739
/organism="Saccharomyces cerevisiae"
/mol_type="genomic DNA"
/strain="W303-1A"
/db_xref="taxon:4932"
/chromosome="XVI"
/gene="TFIIIA"
/note="zinc finger containing protein"
/codon_start=1
/product="transcription factor IIAA"
/protein_id="AAB08014.1"
/db_xref="GI:172902"
/translation="MGKRVLNNEGMDLAELKQETIPISSESESLNSTSRSSSN
RKPTPCDIDGCDKAPTRPSILTEHDLSTHGLRACQCKKSPFKSHLEHLYTH
SDTPKQCSYCGKGVATTRQOLRHEVTHKSTICEEGCNLRFYKAPQAPHLSTH
HLTDPCKKSPQRPYRLNHI SKHIDPEVEMPYOCTFGGCKEPIWGLQSHIND
HKLCPKCSKEPCVGENLOHMIHIDSLVTRMCHICPMSPFRKIDLTHYSI
HTEBDIPLEKTKISDIOQVODHGVQLNSHNSODEKISNRLRKRLTENNV
FLQNEVDIEKRLSEGENLNLITVGNKVCYFNKSGRTFKTEKYEKHIDKHVH
ELKTILOKEKNTLVQNMKEPFIQKETGSAGDK"

ORIGIN

Query Match 17.5%; Score 217.4; DB 8; Length 1739;

Best Local Similarity 59.4%; Pred. No. 1.3e-25;

Matches 409; Conservative 0; Mismatches 271; Indels 9; Gaps 2;

62 GTCCAAAAAGTATTTTGCATATGAAAGGTTGATTAAGCCTATTAATCGACATCAT 121

367 GGCAGAAAGCATATTTCTGTGATATGATGGCTGTGATTAAGCATTTACAGACCTTCAA 426

122 TATTAGAGCAACATTTAAGAACCCAGTAATATGACCGTATTAATGTACAGTGAGC 181

427 TTTTGACTGAACCAATTAAGGTATCATCAGGGTTTA--AGAGCATTTCACTGTGATA 483

182 ATTGTGATTAAGCATTTTTCAGAAATTCATATTGAAACATATTTGATCATCTCCG 241

484 AGGTGCAAAATCTCTGTATTAAGAGTACTTAAAGAACACTGTATAGCATTTCTG 543

242 AAAAAAACCATTCATTTGTCAGTGTGTGTAAGGGTTAATTCGCAACACTTGA 301

544 ATACGAAACATTCATATGTTCTTATTTGTGAAAGAGAGAGACCTCGCAGCACTGA 603

302 AAAGACATGAATCAATCCATATCAAAATGATCAATTTAATGTACATTTGAAATTTGCAAG 361

604 AGGACACAGAAAGTACATACCAATCTTCAATTTGTCCAGAGAGAGATGCAACTCC 663

362 CATTTATTAACATCATCTTTAAGACATCATATATTTATCTGTTCAATGAAAAACATTA 421

664 GATTTCTAAGCATTCACATTAAGGCGCATATTTTATCTGTCTATTATTAATTAATTA 723

422 CGTGAACATATGTAATTAAGTTTCACTGCACTTCAAAATTAAGCAACATTAATTA 481

724 CCGTCCACACTGCAATTAAGCTTTCAAGAGGCGCATATAGGCTTAAGAAATCACTTTCTA 783

AUTHORS Archambault, J., Schappert, K.T. and Friesen, J.D.
 TITLE A suppressor of an RNA polymerase II mutation of *Saccharomyces cerevisiae* encodes a subunit common to RNA polymerases I, II, and III
 JOURNAL Mol. Cell. Biol. 10 (12), 6123-6131 (1990)
 MEDLINE 91061718
 PUBLISHED 2247052
 COMMENT Original source text: *S. cerevisiae* single-copy plasmid DNA, clone 2.
 Draft entry and computer-readable sequence for [unpublished (1990)] kindly submitted by J.D. Friesen, 01-MAY-1990
 The Hospital for Sick Children
 555 University Avenue
 Toronto, Ontario, CANADA M5G 1X8
 e-mail: jfmesick@sickkids.toronto.edu.
 Location/Qualifiers

FEATURES
 source 1..2133
 /organism="Saccharomyces cerevisiae"
 /mol_type="genomic DNA"
 /db_xref="taxon:4932"
 join(1026..1045,1122..1569)
 /note="RNA polymerase II sixth subunit (RP026)"
 /codon_start=1
 /protein_id="AAA34989.1"
 /db_xref="GI:172453"
 /translation="MSDYEAFAFDNGENFEDVEHFDSETEYEKPKQDETTDAN
 GRTIVGNGPDPFOHIOIRKTLKEKAIIPDORATTPYMKRYEAKRIIGRAIOIS
 NNAIPVDEEGEDPLRIKMKLEKIKPLIVIRRLIPDSFPMSEVELIYDL"
 <1026..1045
 /note="RNA polymerase II sixth subunit (RP026)"
 /number=1
 1046..1121
 /note="RP026 intron A"
 1122..1569
 /note="RNA polymerase II sixth subunit (RP026)"
 /number=2

ORIGIN Chromosome XVI.

Query Match 16.1%; Score 199.2; DB 8; Length 2133;
 Best Local Similarity 61.4%; Pred. No. 1.1e-22;
 Matches 340; Conservative 0; Mismatches 208; Indels 6; Gaps 1;

171 TACAGTGAAGCATTTGTATTAAGCATTTTTCAGAAAATACATTTGGAAACATATTGT 230
 556 TTCAAGTGATAGTGTGCAAAATCCCTTCTTAAAGAGTCACTTAGAGACACTGTGA 497
 231 ATCACATTCGGAATAAAACCATTCATTTGTGAGTGTGTGAAGGGGTTAATTCCTG 290
 496 TACGCAATTCGTATACGAACCAATTCMAATGTTCTTAATGTGAAAAGAGTACGACTCG 437
 291 ACAACACTGAAAAGACATGAATACCCGATCAAAATGATTTAAATGTATCAATTTGAAA 350
 436 CCGCACTGAAGCGACGAAGTAAAGCATACCAATTTTCATTTGTTCAGAGAGAG 377
 351 TTGTCAAGAAGCATTTTATTAACATCTTTAAGACATCATATTAATTCGTTTCATGA 410
 376 ATGCAACCTCCGATTTTACAGCATCCAAATTAAGGCAATATTTATTCGTTTCATTT 317
 411 AAAACATTACGTGTAAACATGTATTAAGTTTCACTCGACCTTCAAAATTAGACA 470
 316 ACATTAACCTACCTGTCCACACTGCACAAATTAAGCTTCAGAGGCCATATAGGCTAAGAA 257
 471 ACATTAATTAACATCA-----TGGTGATCTCTGCTTATCATATGATCATATCCTG 524
 256 TGCATTTTCAACATCAAGATCTGAGTAGAAAATCCCTTACCAATGATCTTTGCTCG 197
 525 TTGTTTAAAAATTTCCAACTGTGCTAGTATTAACATTTCAATTAATAACAACTGCATCC 584
 196 TTGTTGCAAGAAGTTTCGATATGTATGTCACATTTGCATATGCAATTAATAAAGATCATCC 137
 585 AAAACTTAATGTCTTAAATGTGTAAAGGTTGTGTGGAAAAAAGTTTATCTTCACA 644

Db 136 TAAATTAAGTCTTATTTGTAGCAACATGTGTGGGGAAAAATGTTTAAATGCA 77
 Oy 645 TATGTTAAGTCAATGATGATTTCTACCATGATCAAAATATGACTTGTGATTAATGTGATGT 704
 Db 76 CATGATTTATCTAGACGACTCACTAGTAAACAAAATTTGGAAGTCTATATATGTCCTGA 17
 Oy 705 GGGGAAATTTGCAA 718
 Db 16 TATGCTTTTCTA 3

RESULT 13
 CDS06ING 797 bp DNA linear STS 30-NOV-2001
 LOCUS T7 end of clone AS0AA022B07 of library AS0AA from strain CLIB 533
 DEFINITION of *Saccharomyces bayanus*, sequence tagged site.
 ACCESSION AL400514
 VERSION AL400514.1 GI:12156635
 KEYWORDS STS.
 ORGANISM *Saccharomyces bayanus*
 SOURCE *Saccharomyces bayanus*
 Eukaryota; Fungi; Ascomycota; *Saccharomycotina*; *Saccharomycetes*;
Saccharomycetales; *Saccharomycetaceae*; *Saccharomyces*.

REFERENCE 1 (bases 1 to 797)
 Souleir, J.L., Aigle, M., Attiguenave, F., Blandin, G.,
 Bolotin-Fukuhara, M., Bon, E., Broctier, P., Casaregola, S.,
 de-Montigny, D., Dujon, B., Durrens, P., Leplingle, A., Llorente, B.,
 Malpertuy, A., Neuvéglise, C., Ozier-Kalogeropoulos, O., Pottier, S.,
 Saurin, W., Tekala, F., Toffano-Nicche, C., Wesolowski-Louvel, M.,
 Wincker, P. and Weissenbach, J.
 Genomic exploration of the hemiascomycetous yeasts: 1. A set of
 yeast species for molecular evolution studies
 FEBS Lett. 487 (1), 3-12 (2000)

TITLE 2 (bases 1 to 797)
 JOURNAL Bon, E., Neuvéglise, C., Casaregola, S., Attiguenave, F., Wincker, P.,
 MEDLINE 20584711
 PUBLISHED 11152876
 REFERENCE 3 (bases 1 to 797)
 JOURNAL 11152880
 MEDLINE 20584715
 PUBLISHED 11152880

TITLE Direct Submission
 JOURNAL Submitted (07-SEP-2000) Genoscope - Centre National de Séquençage,
 2 rue Gaston Crémieux, CP 5706, 91057 EVRY cedex, FRANCE. (E-mail :
 seque@genoscope.cns.fr - Web : www.genoscope.cns.fr)

COMMENT This STS is part of a random genomic sequencing program of thirteen
 yeast species: *Saccharomyces bayanus* var. *uvaurum*, *Saccharomyces*
exiguus, *Saccharomyces servazzii*, *Zygosaccharomyces rouxii*,
Saccharomyces kluyveri, *Kluyveromyces thermotolerans*, *Kluyveromyces*
lactis var. *lactis*, *Kluyveromyces marxianus* var. *marxianus*, *Pichia*
angusta, *Debaryomyces hansenii* var. *hansenii*, *Pichia sobolofphila*,
Candida tropicalis and *Yarrowia lipolytica*. Genomic inserts of 3 to
 5 kb were prepared and both extremities were sequenced. See
 keywords for description of this sequence and for the sequence of
 the other extremity of this insert.

FEATURES
 source 1..797
 /organism="Saccharomyces bayanus"
 /mol_type="genomic DNA"
 /strain="CLIB 533"
 /variety="uvaurum"
 /db_xref="taxon:4931"
 /clone="AS0AA022B07"
 /clone_11b="AS0AA"
 /note="end : 797"
 <2..>796
 /note="similar to *Saccharomyces cerevisiae* ORF YPR186c [
 PZP1 ; TP11A (transcription initiation factor)]"
 /evidence="not_experimental"

ORIGIN

Query Match 15.6%; Score 193; DB 11; Length 797;
Best Local Similarity 59.3%; Pred. No. 1.4e-21;
Matches 347; Conservative 1; Mismatches 231; Indels 6; Gaps 1;

QY 178 GACGATTTGATTAAGCAATTTTTCAGAAATCAATTTGGAAACACATATTGATCAT 237
DB 83 GACAAATTTGCAAAATCCTTTGTTAAABAAGATCATTAAGAGGACCTGTTTGGCAT 142
QY 238 TCCGAAAAAAACCATTTCCATTTGTCAGTGTGTGTAAAGGGTTAATTTCGACAAAC 297
DB 143 TCTGATACGAAGGCATTTCCATTTCTTATTGTGGCAAGGGGGTGAACAATCCAGACAA 202
QY 298 TTGAAGAAGATGAATACCCATACCAATGATTAATGATGATTTGAAATTTGCA 357
DB 203 TTGAAGGCCAGAAAGTTACACAAATCTTTGTTGTCCGGAAGTGAATGTGAT 262
QY 358 GAAGCATTTTAAACATCAATCTTTTAAAGATCATATATTATCTGTTCAAAAAACA 417
DB 263 CTCGATTTTAAAGACACCTCAATTAAGGGCTCATTTCTATCTGTCCACCTAGACAA 322
QY 418 TTAACGTGAACAATGATTAATAAGTTTCACTGACCTTCAAAATTAAGCAACAATAA 477
DB 323 TTAACCTGTCCAAATGATTAATAAATTTTCAAGAACCTTACAGGCTTAAGGAATCATATC 382
QY 478 TTAACATCATATGT-----GGATCTCCCTGCTTATCATATGATCATCTGTTGTTT 531
DB 383 TCTAAACATCATATCTCTGAAGGTGAATATCCGATCAATGATCTTTCGCTGTGTGT 442
QY 532 AAAATTTTCCAACTTGTCAGTATTAATTTCAATTAATAACAATGATCAAAACT 591
DB 443 ACAGATACCGTATATGTGTACAGTGTGATCATCATTAACAAATGACATCTTAATTTG 502
QY 592 AAATGCTTAATGTGTGAAGGTGTGTGGAAAAAGTTTATCTTACATATGTTA 651
DB 503 AAATGCCCCATTGGCAACAAGCCTTGTGTAGGGGAAATGTCTCAAAATGATATGATC 562
QY 652 AGTCATGATGATTTTACATGATCAAAATATGACCTTGTGATTTATGTGATGTGGGAAA 711
DB 563 ATTCAATGATGAATCTTGTGATCAAGTGAAGTGAAGTCAATTTGTTCTAATATGATCT 622
QY 712 TTGCAAGAAAAATGATTAATGATTAATTAATTAATTAATTAATTAATTAATTAAT 756
DB 623 TTCTTGAGAAAGCAAGCTCTTACTTACTTACCAAGTACGAT 667

RESULT 14
AE016818 01

WPCOMMENT

Sequence split into 16 fragments LOCUS AE016818 Accession AE016818

Fragment Name	Begin	End
AE016818_00	1	110000
AE016818_01	100001	210000
AE016818_02	200001	310000
AE016818_03	300001	410000
AE016818_04	400001	510000
AE016818_05	500001	610000
AE016818_06	600001	710000
AE016818_07	700001	810000
AE016818_08	800001	910000
AE016818_09	900001	1010000
AE016818_10	1000001	1110000
AE016818_11	1100001	1210000
AE016818_12	1200001	1310000
AE016818_13	1300001	1410000
AE016818_14	1400001	1510000
AE016818_15	1500001	1519138

Continuation (2 of 16) of AE016818 from base 100001 (AE016818 Eremothecium gossypii ATCC

Query Match 9.7%; Score 120.6; DB 8; Length 110000;
Best Local Similarity 50.1%; Pred. No. 1.9e-10;
Matches 386; Conservative 0; Mismatches 374; Indels 11; Gaps 3;

QY 25 TCGATATCATCTTTAATATCTTCTCTTCAATCAAGTCCAAAAAGTATATTGACA 84
DB 17349 TCGTTTTCAGATTTGAATGTCACGCGGTGAGACACCGCGGCAAAACGATTTCTGCGAC 17408
QY 85 TATGAAGGTGTGATTAAGCTTAATTCGACATCATTTATTAGACAAACA-TTAAAGAAC 143
DB 17409 TACGAGGGCTGCTCAAAAGGGTTTCAAGCGCGCTCCCTGTGTGACGGAGCACACGAGACG 17468
QY 144 CCACAGTAATGATCGACCGTATTAATGTACAGTGAACGATTTGATTAAGCATTTTTCAG 203
DB 17469 GCGCACACAGGGCATTCAGGGGATTCACAGTGC---GAGCATGTGCGGCGCGGTTTACAA 17524
QY 204 AAAATCACTTTGGAAGAACATATTTGATTCACATTCGAAAAAAACCAATTCATTTGTC 263
DB 17525 GAAGTGCACCTTGAGACCGGACCTGTTCTGCACTTGTAGACAAAGCGTTTACGCTGCAC 17584
QY 264 AGTGTGTGTTAAAGGGGTTAATTTCTGACAAACATTTGAAAGACATGAAATCACCCATAC 323
DB 17585 GGTATGTGCGAAGGGGTGACACCGCGGACGACATGAGCGGCGACAGATTCACACAC 17644
QY 324 AAAATCATTTAAATGTACATTTGAAATTTGTCAAGACATTTTATTAATCATCATCTTT 383
DB 17645 AAAGTGTTCAGATGCGCCGACAGAGGCTGCGGCAAGGCTTCAAGACACCGCAGTT 17704
QY 384 AAGACATCATATATTTATCTGTTATGAAAAAAACATTAAGGTGAACATATTAATTAAGT 443
DB 17705 GCGTTCGACAGTGTGCGGCTGACAGCAAAAGTGAAGTGTGACACATGTACAAAGCG 17764
QY 444 TTCACTGACCTTCAAAATTAAGCAACAATTAATTAACATCATGTTGATCTCC--- 500
DB 17765 GTTCCAAAGGCGCATACCGTTTGAACACACATGCGCAAGATTCAGGGCTGATGCA 17824
QY 501 ---TGTATCAATGTGATCATCTGTTTAAATTTTCAAACTTGTGATCTGAT 557
DB 17825 GTTCCGATCAATGACAGCAAGCGCGGTGCGTGAAGTGTGAGACGCGTGTGCGCT 17884
QY 558 ACAATTTCAATTAACAACTGATTCAAACATTAATGCTTAATGTGTGAAGTTG 617
DB 17885 GCAGCAGACCTTCAACAGGACCAACCGAAGCTACATGCGGGGTGTGCGGAAAGCTGTG 17944
QY 618 TGTGGAAGAAAAAGTTTATCTTCAATATGTTAAGTCAATGATGATTTTACATGATCA 677
DB 17945 CTTGGGGAGAGCGGACTGCAATGCAATGCTGTGACAGAGTCCGGGTGATTTAA 18004
QY 678 AATATGACTTGTATTTGTGATGTGGGAAATTTGCAAGAAATGAATTAAGTTGA 737
DB 18005 AAACGTGAAGTGTCTGTGTGTGACGACACACCTACGCGAAGATGCGCATGCTTGC 18064
QY 738 ACATTATTAATATCTTCAATGATGATTAATATCCCTGATGATTTATTAAGGA 788
DB 18065 GCATTATTAAGAAACCCACGATGATTCATCCCAAGAAATGATTAAGGA 18115

RESULT 15

AC138126/c

LOCUS AC138126 169063 bp DNA linear PRI 17-DEC-2002

DEFINITION Homo sapiens chromosome 19 clone RP11-274A19, complete sequence.

AC138126

VERSION AC138126.1 GI:27151357

KEYWORDS

HTG.

SOURCE

ORGANISM

Homo sapiens (human)

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.

1 (bases 1 to 169063)

DOE Joint Genome Institute and Stanford Human Genome Center.

Direct Submission

Unpublished

Submitted (17-DEC-2002) DOE Joint Genome Institute, 2800 Mitchell

COMMENT

Drive, Walnut Creek, CA 94598, USA
Draft Sequence Produced by DOE Joint Genome Institute
www.jgi.doe.gov
Finishing Completed at Stanford Human Genome Center
www-shgc.stanford.edu
Quality: Phrap Quality >=40 100% of Sequence;
Estimated Total Number of Errors is 0.

FEATURES

source

1..169063
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="19"
/clone="RP11-274A19"

ORIGIN

Query Match 8 2%; Score 101; DB 9; Length 169063;
Best Local Similarity 50.8%; Pred. No. 2.4e-07;
Matches 298; Conservative 0; Mismatches 280; Indels 9; Gaps 2;

QY 7 GAAAGTGACGAAACCAATCGATATCATCTTTAATATCTTCTTCATCACGTCCTCC 66
DB 80134 GAATGGGCAAGCTTTTAAACGAGTCCTCAATCCTTACTACACATAGAGAAATCTACT 80075
QY 67 AAAAGTATATTGCACATATGAGGGGTGTATAAGCCTATATGACCATCATATTATTA 126
DB 80074 GGAGGAAATCCTACAAATGTGAAGATGTGGCAAAAGCTTCTATCATGATCTCAAAACTT 80015
QY 127 GAGCAACATTTAAGAACCAACGATATGATGACCGTATTAATGTATGATGACGATGCT 186
DB 80014 ACTGAACATTAAGAAATTTATCTGAGAGAAACCTTACACAGT-----GAGAGATGT 79961
QY 187 GATTAAGCATTTTTCAGAAATTCACATTTGAAACATATTTGATCATCTCCGAAAAA 246
DB 79960 GGCAAGCCTTTAACCATCTCTACATCTGCTACACATAGGTAATTCATCTGAGAG 79901
QY 247 AACCATTCATTTGTCAGTGTGTGTAAAGGGGTTAATTCGACACACTTGAAAAAGA 306
DB 79900 AACCTTACCAATGTGAAGAAATGTGTAAGCCTTTAACAGTCTCAACCTTACTAGA 79841
QY 307 CATGAATACCCATACCAAGATCATTTAATGTACATTTGAAATTTGTCAAGAAGCATTT 366
DB 79840 CATAGAGAAATTCATACCTGAGAGAAACCTTACCAATGTGAAAAAATGTGCAAAAGCTTTT 79781
QY 367 TATTAACATCAATCTTTAA--GACATCATATATTATCTGTCATGAAAAAACATTACG 423
DB 79780 AACCACTCTCAAACTTACTGACATTAAGAAATTCATCTGCTGAGAACTTACAAA 79721
QY 424 TGTAAACAATGTATTAAGTTTCACTGACCTTCAAAATTAAGCACACATTAATTAATA 483
DB 79720 CCTAAAGATGTAAAGATTTTGAACACCTTCAAAGTTTCTTAAACATTAAGAAAT 79661
QY 484 CATCATGTGATCTCTGCTTATCATGTGATCATCTGTTGTTTAAATTTCCAA 543
DB 79660 TATGCTGAGTGAAGAAATCTTGAAGATGTGAAGATGTACAAAACCTTTAAAGTTGTAC 79601
QY 544 ACTTGGTCAGTATTAATTAATTAATAAACAACGATCCAAAACCT 590
DB 79600 ACTTGATTTGTGATTAAGATTAATTCATTAATAAAAAAACCTTACAAAGT 79554

Search completed: February 9, 2005, 05:36:23
Job time : 5295.91 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 8, 2005, 22:43:04 ; Search time 642.278 Seconds
(without alignments)
11419.610 Million cell updates/sec

Title: US-09-831-804-2
Perfect score: 1239
Sequence: 1 atgagctgaagacgacgaac.....catcagtgattctctgataa 1239

Scoring table: IDENTITY_NTC
Gapop 10.0 , Gapext 1.0

Searched: 4390206 seqs, 2959870667 residues

Total number of hits satisfying chosen parameters: 8780412

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_16Dec04:*
1: geneseqn1980s:*
2: geneseqn1990s:*
3: geneseqn2000s:*
4: geneseqn2001as:*
5: geneseqn2001bs:*
6: geneseqn2002as:*
7: geneseqn2002bs:*
8: geneseqn2003as:*
9: geneseqn2003bs:*
10: geneseqn2003cs:*
11: geneseqn2003ds:*
12: geneseqn2004as:*
13: geneseqn2004bs:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	1239	100.0	2060	3	AAAI5398	AAAI5398 DNA encod
2	1235.8	99.7	1239	6	ABZ32205	ABZ32205 Candida a
3	99.4	8.0	2132	10	ADAA53158	ADAA53158 Human cod
4	96.8	7.8	1705	9	ACA98970	ACA98970 cDNA enco
5	96.6	7.8	2320	10	ACA56456	ACA56456 Human sig
6	96.6	7.8	2320	10	AD156252	AD156252 Human pol
7	95.2	7.7	4563	5	AA931317	AA931317 DNA encod
8	95	7.7	981	12	ADNO4717	ADNO4717 Antipsoi
9	94.6	7.6	2597	8	AAI51569	AAI51569 Human nuc
10	92.4	7.5	3639	5	AA864586	AA864586 DNA encod
11	92	7.4	994	12	ACH87568	ACH87568 Human gen
12	92	7.4	2873	13	ADRI4368	ADRI4368 Human NF-
13	91.8	7.4	1533	12	ADN99064	ADN99064 Novel hum
14	91.8	7.4	1533	12	ADDO0633	ADDO0633 Novel hum
15	91.8	7.4	2110	10	ADDB3624	ADDB3624 Human cDN
16	91.8	7.4	2760	10	ADCS6695	ADCS6695 Human mac
17	91.8	7.4	3078	10	ADBE2468	ADBE2468 Human cDN
18	91	7.3	3755	13	ADRO7315	ADRO7315 Full leng
19	91	7.3	4209	13	ADRO8223	ADRO8223 Full leng
20	90.4	7.3	2026	8	AA55863	AA55863 Human nuc

21	90.4	7.3	2114	10	ADA53124	ADA53124 Human cod
22	90.2	7.3	2230	10	ADA52931	ADA52931 Human cod
23	89.8	7.2	590	12	ACH80103	ACH80103 Human gen
24	89.2	7.2	2064	10	ADCB30762	ADCB30762 Human nov
25	89.2	7.2	2221	11	ADM01344	ADM01344 Human cDN
26	89.2	7.2	2622	5	AA668872	AA668872 DNA encod
27	89.2	7.2	2729	4	AAH16178	AAH16178 Human cDN
28	89.2	7.2	3839	6	ABK83826	ABK83826 Human cDN
29	89.2	7.2	3839	12	ADNO4207	ADNO4207 Antipsoi
30	89.2	7.2	3839	13	ADRB25049	ADRB25049 Breast ca
31	89	7.2	1991	13	ADSO9824	ADSO9824 Human the
32	89	7.2	2298	4	AAH16608	AAH16608 Human cDN
33	89	7.2	2299	13	ADSO9825	ADSO9825 Human the
34	89	7.2	2905	5	AA666143	AA666143 DNA encod
35	89	7.2	3003	12	ADNO6033	ADNO6033 Antipsoi
36	89	7.2	3020	5	AA92560	AA92560 DNA encod
37	89	7.2	3502	4	AA157845	AA157845 Human pol
38	88.6	7.2	1335	12	ADN98906	ADN98906 Novel hum
39	88.6	7.2	1335	12	ADDO0475	ADDO0475 Novel hum
40	88.6	7.2	1466	12	ACH91879	ACH91879 Human gen
41	88.6	7.2	1875	10	AD160591	AD160591 Secreted
42	88.6	7.2	2597	8	AD55885	AD55885 Human nuc
43	88.6	7.2	4227	9	ACA98938	ACA98938 cDNA enco
44	88.4	7.1	1410	12	ACH87337	ACH87337 Human gen
45	88.4	7.1	1952	10	ADC58104	ADC58104 Zinc fing

ALIGNMENTS

RESULT 1	
ID	AAAI5398 standard; DNA; 2060 BP.
AC	AAAI5398;
XX	
XX	
XX	04-SEP-2000 (first entry)
DE	DNA encoding a transcription factor designated CATP11A.
XX	
XX	Transcription factor; CATP11A; DNA-binding protein;
KW	ribosomal RNA 5S gene; fungal infection; ss.
XX	
OS	Candida albicans.
XX	
FH	Key
FT	CDS
FT	Location/Qualifiers
FT	720..1958
FT	/*tag= a
FT	/catal_except= (pos: 1296..1298, aa: Ser)
FT	/catal_except= (pos: 1734..1736, aa: Ser)
XX	
XX	WO200028037-A1.
XX	
PD	18-MAY-2000.
XX	
PF	09-NOV-1999; 99WO-FR002739.
XX	
PR	10-NOV-1998; 98PR-00014147.
XX	
PA	(HMRI) HOECHST MARION ROUSSEL.
XX	
PI	Bordon-Pallier F, Camier S, Sentenac A;
XX	
DR	WPI; 2000-376549/32.
DR	P-PSDB; AAY93316.
XX	
PT	New nucleic acid encoding Candida albicans transcription factor, useful
PT	e.g. in screening for antifungal agents and for immunization.
XX	
PS	Claim 4; Page 32-33; 45DP; French.
XX	
CC	The present sequence encodes a Candida albicans transcription factor, is designated CATP11A. The polypeptide is a DNA-binding protein, which is

CC involved in initiating transcription of the ribosomal RNA 5S gene. The
CC polynucleotide is used to screen for its specific inhibitors, potentially
CC useful as antifungal agents, to raise an antibody response that is
CC protective against fungal infection and to raise antibodies. Such
CC antibodies, as well as the polypeptides and polynucleotides are used in
CC compositions for diagnosing and treating fungal infections, e.g. by
CC detecting polymorphisms and mutations

XX Sequence 2060 BP; 726 A; 315 C; 354 G; 665 T; 0 U; 0 Other;

Query Match 100.0%; Score 1239; DB 3; Length 2060;
Best Local Similarity 100.0%; Pred. No. 3.5e-234;
Matches 1239; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
QY 1 ATGAGTGAAGTGAACCAATGATATCATCTTAAATCTCTCTTCTGATCA 60
DB 720 ATGAGTGAAGTGAACCAATGATATCATCTTAAATCTCTCTTCTGATCA 779
QY 61 CGTCCCAAAAGTATTTTGACATATGAAAGGTGATPAAAGCTATPAAATGACCATCA 120
DB 780 CGTCCCAAAAGTATTTTGACATATGAAAGGTGATPAAAGCTATPAAATGACCATCA 839
QY 121 TTATTAAGACATCTTAAGAACCCACATPAAATGACGACCGTAAATATGTACAGTGAC 180
DB 840 TTATTAAGACATCTTAAGAACCCACATPAAATGACGACCGTAAATATGTACAGTGAC 899
QY 181 GATGTGATPAAAGCATTTTTCAGAAATGACATTTTGAACACATATTTGATCATCTCC 240
DB 900 GATGTGATPAAAGCATTTTTCAGAAATGACATTTTGAACACATATTTGATCATCTCC 959
QY 241 GAAAAAAACATTCATTTGTCAGTGTGTAAAGGGGTAAATCTCGACAAACATTTG 300
DB 960 GAAAAAAACATTCATTTGTCAGTGTGTAAAGGGGTAAATCTCGACAAACATTTG 1019
QY 301 AAAAGCATGAAGTACCCATCAAAAGTATTAATGATPAAATGATPAAATTTGTCAAGA 360
DB 1020 AAAAGCATGAAGTACCCATCAAAAGTATTAATGATPAAATGATPAAATTTGTCAAGA 1079
QY 361 GCATTTTAAACATCATCTTTAAAGACATCATATATATCTGTTGATGAAAAACATTA 420
DB 1080 GCATTTTAAACATCATCTTTAAAGACATCATATATATCTGTTGATGAAAAACATTA 1139
QY 421 ACGTGAACATGATPAAAGTTTCACTCGACCTTCAAAATAGACAAACATPAAATTA 480
DB 1140 ACGTGAACATGATPAAAGTTTCACTCGACCTTCAAAATAGACAAACATPAAATTA 1199
QY 481 AAAACATGATGATGATCTCTGTTATCAATGATCATCTCGTGTGTTTAAATTTTC 540
DB 1200 AAAACATGATGATGATCTCTGTTATCAATGATCATCTCGTGTGTTTAAATTTTC 1259
QY 541 CAAACTTGGTCAATTTTAAATTTCAATPAAACATGATCATCAAACTTAAATGTCCT 600
DB 1260 CAAACTTGGTCAATTTTAAATTTCAATPAAACATGATCATCAAACTTAAATGTCCT 1319
QY 601 AAATGTGTAAGTGTGTGGAAAAAGGTTTATCTTCAATATGTTAAGTCATGAT 660
DB 1320 AAATGTGTAAGTGTGTGGAAAAAGGTTTATCTTCAATATGTTAAGTCATGAT 1379
QY 661 GATTTCAACATGATCAAAATATGAGATCTTGATATTTGATGTGGGAAATTTGCAAG 720
DB 1380 GATTTCAACATGATCAAAATATGAGATCTTGATATTTGATGTGGGAAATTTGCAAG 1439
QY 721 AAAAATGAATTAATTAATATCTTCATGATGATPAAATGATGATGATTTA 780
DB 1440 AAAAATGAATTAATTAATATCTTCATGATGATGATPAAATGATGATGATTTA 1499
QY 781 TTTAAAGAAATGAAATGAAATTTAGAACTTATTAATGATCATGATGAAATTAAT 840
DB 1500 TTTAAAGAAATGAAATGAAATTTAGAACTTATTAATGATCATGATGAAATTAAT 1559
QY 841 AATTTGATGATTTAGAAACAGAAATTTAAAGTGAAGAGATGAAGATGAAGAA 900
DB 1560 AATTTGATGATTTAGAAACAGAAATTTAAAGTGAAGAGATGAAGATGAAGAA 1619
```

```
QY 901 GATAGCTAGATGAAAAAGAGTATGTATGATCAGATCATATGTCAGCTCAAGATCA 960
DB 1620 GATAGCTAGATGAAAAAGAGTATGTATGATCAGATCATATGTCAGCTCAAGATCA 1679
QY 961 ATAAATCATTTTACTGCTTTCTTTGGAAGTTCAAAAGATGTTCTTAACTTATTCGAT 1020
DB 1680 ATAAATCATTTTACTGCTTTCTTTGGAAGTTCAAAAGATGTTCTTAACTTATTCGAT 1739
QY 1021 AGTGGAGAGATCAATTTGCTTAAGATATATTTGATGATGATGTTTCTGAGAAATP 1080
DB 1740 AGTGGAGAGATCAATTTGCTTAAGATATATTTGATGATGATGTTTCTGAGAAATP 1799
QY 1081 GATTTACGTGACATTTGAAATGCGATGATGATPAAATTTTAAAGAAATGATCATTTCTTA 1140
DB 1800 GATTTACGTGACATTTGAAATGCGATGATGATPAAATTTTAAAGAAATGATCATTTCTTA 1859
QY 1141 AATAGTATPAAAGAAAGAACTCCAGAGGTGAACATTTGTTAAAAAGCCGAGATG 1200
DB 1860 AATAGTATPAAAGAAAGAACTCCAGAGGTGAACATTTGTTAAAAAGCCGAGATG 1919
QY 1201 GATTTATGCGAAATGAACATCAGTGAATTTCTGATPAA 1239
DB 1920 GATTTATGCGAAATGAACATCAGTGAATTTCTGATPAA 1958
```

```
RESULT 2
AB32206
ID AB32206 standard; DNA; 1239 BP.
XX
AC AB32206;
XX
DT 30-JAN-2003 (first entry)
XX
DE Candida albicans essential gene SEQ ID NO 6493.
XX
KW Fungus; yeast; tetracyclin; promoter; GRACE strain; biosynthesis;
KW signal transduction; DNA replication; cell division; growth;
KW proliferation; Candida albicans; fungicide; antifungal; gene; ss.
XX
OS Candida albicans.
XX
PN WO200253728-A2.
XX
PD 11-JUL-2002.
XX
PP 26-DEC-2001; 2001MO-US049486.
XX
PR 29-DEC-2000; 2000US-0259128P.
PR 20-FEB-2001; 2001US-0079202A.
PR 22-AUG-2001; 2001US-0314050P.
XX
PA (ELITR-) ELITRA PHARM INC.
XX
PI Roemer T, Jiang B, Boone C, Buesey H, Ohlsen KL;
XX
DR WPI: 2002-566694/60.
DR P-PsDB; ABR73656.
XX
PT Constructing strains for identifying gene products as effective targets
PT for therapeutic intervention, by inactivating in the strain one allele of
PT a gene and placing other allele of the gene under conditional expression.
XX
PS Claim 37; SEQ ID NO 6493; 167bp + Sequence Listing; English.
XX
CC The invention relates to constructing (M1) a strain of diploid fungal
CC cells in which both alleles of a gene are modified, comprising modifying
CC one allele by insertion or replacement by a cassette having an
CC expressible selectable marker and modifying other allele by
CC recombination of a promoter replacement fragment with a heterologous
CC promoter, so that expression of the second allele is regulated by the
CC promoter. (M1) is useful for constructing a strain of diploid fungal
CC cells in which both alleles of a gene are modified. The diploid fungal
```


PT which the gene is involved, or as target molecules for gene therapy.
XX
XX Claim 1; SEQ ID NO 726; 205pp; English.
XX
XX The present invention relates to novel human secretory or membrane
CC proteins (ADA54072-ADA55710) and their coding sequences (ADA5433-
CC ADA54077). The coding sequences are useful in the gene therapy of
CC diseases caused by abnormalities of the proteins, e.g. cancer,
CC inflammatory diseases, osteoporosis or neurological disease.
XX
SQ Sequence 2132 BP; 833 A; 423 C; 347 G; 529 T; 0 U; 0 Other;

Query Match 8.0%; Score 99.4; DB 10; Length 2132;
Best Local Similarity 50.6%; Pred. No. 4.7e-10;
Matches 297; Conservative 0; Mismatches 281; Indels 9; Gaps 2;

QY 7 GAAAGTGAAGCAACCAATCGATATCATCTTTAATCTCTCTCTTCATCAGCTCC 66
DB 1384 GAATGTGGCAAGCTTTTAACCGCTCTCATCTTACTACATAGAGATTCATCT 1443
QY 67 AAAAGTATATTTGACATATGAAAGGATGATTAAGCTATATCGACCATATTATTA 126
DB 1444 GAGAGAAATCCACAAATGTGAAGATGTGCAAGCTTCTATGATCTCAAAACTT 1503
QY 127 GAGCAATTTAAGAACCAAGTAATGATCGACGTATATATGATGACGATGT 186
DB 1504 ACTGAACATTAAGAAATTTCACTGAGAGAAACCTTACACATGT-----GAAGAATGT 1557
QY 187 GATTAAGCATTTTTCGAAATATGACATTTGAAACATATTTGATACATTCGGAATA 246
DB 1558 GGCAGAACCTTTAACCATCTCTACACCTTGCTACATTAAGTATTTCACTGAGAG 1617
QY 247 AAACCATTCATTTGTCAGTGTGTGTTAAAGGGGTTAATCTTCGACACATTTGAAAGA 306
DB 1618 AAACCTTACCAATGTGAAGATGTGTTAAAGCTTTAACAGCTTCACACTTACTA 1677
QY 307 CATGAATACCCATACAAAGTCATTTAAATGTCAATTTGAAATTTGCAAGAGCATTT 366
DB 1678 CATTAAGATTCATCTGAGAGAAACCTTACCAATGTGAAATTTGGCAAGCTTTT 1737
QY 367 TATTAACATCATCTTTAA---GACATCATATTTATCTGTCATGAAATTAACATTAAG 423
DB 1738 AACCATCTCTCAAACTTACTGTGACATTAAGAAATTCATCTGTAGAAACTTACAA 1797
QY 424 TGTAAACAATGTAAATTAAGTTTCACTCGACCTTCAAAATTTAGCAACATTAATTA 483
DB 1798 CCGAAAGATGTAAGTATTTTGAACAACCTTCAAACTTTCTAAACATTAAGAAAT 1857
QY 484 CATCATGTGTGATCTCTGCTTATCAATGTATCATCTGTGTTTAAATTTCCAA 543
DB 1858 TATGCTGTGAGAAATCTTGAATGTGAAGATGTAAACAAACCTTTAAAGTTGTAC 1917
QY 544 ACTGTGTGATTTACATTTCAATTAACAAACATCCAAACT 590
DB 1918 ACTGATGTGTGATTAATTAATTCATTAACAAACCTTAAGT 1964

RESULT 4
ACA98970
ID ACA98970 standard; cDNA; 1705 BP.
XX
XX ACA98970;
XX
DT 25-JUL-2003 (first entry)
XX
DE cDNA encoding human nucleic acid-associated protein (NAAP) #51.
XX
XX Human; nucleic acid-associated protein; cytosolic; antiarteriosclerotic;
KM anticonvulsant; nootropic; neuroprotective; cerebroprotective; anti-HIV;
KM anti-allergic; anti-inflammatory; thymometric; gene therapy;
KM cell proliferative disorder; cancer; atherosclerosis;
KM neurological disorder; epilepsy; Huntington's disease; stroke;
KM immune disorder; inflammatory disorder; AIDS; allergy;

KW developmental disorder; Hypothyroidism; Cushing's syndrome; infection;
KM protein-protein interaction; drug-target interaction;
KW gene expression profile; gene; ss.
XX
XX Homo sapiens.
XX
XX WO2003023003-A2.
XX
XX 20-MAR-2003.
XX
XX 05-SEP-2002; 2002MO-US028540.
XX
XX 07-SEP-2001; 2001US-0317792P.
XX
XX 07-SEP-2001; 2001US-0317912P.
XX
XX 14-SEP-2001; 2001US-0322270P.
XX
XX 21-SEP-2001; 2001US-0324040P.
XX
XX 28-SEP-2001; 2001US-0326732P.
XX
XX 19-OCT-2001; 2001US-0346715P.
XX
XX 25-JAN-2002; 2002US-0351749P.
XX
XX 22-FEB-2002; 2002US-0359498P.
XX
XX (INCY-) INCYTE GENOMICS INC.
XX
XX Tang YT, Jackson JL, Griffin JA, Elliott VS, Forsythe IU;
PI Becha SD, Richardson TW, Lee EA, Sprague WW, Emerling BW;
PI Thangavelu K, Warren BA, Tran UK, Yue H, Xu Y, Yue H, Li JX;
PI Hafalia AJP, Sanjanwala B, Margulis JP, Gervad AE, Lee SY, Ison CH;
PI Baugh MR, Chawla NK, Nguyen DB, Swarnakar A, Zebardjian Y, Shah P;
PI Thornton M, Yao MG, Khan FA, Gandhi AR, Yang J, Kabile AE;
PI Burford N, Ramkumar J;
XX
XX WPI; 2003-313243/30.
XX
XX P-PSDB; AB096722.
XX
XX New human nucleic acid associated proteins (NAAP), useful for diagnosing,
PT creating and preventing diseases or conditions associated with the
PT aberrant NAAP expression e.g. cancer, AIDS, atherosclerosis, epilepsy, or
PT infections.
XX
XX Claim 5, Page 340-341; 345pp; English.
XX
XX The invention describes a novel human isolated nucleic acid-associated
CC polypeptide (NAAP). The polypeptides and polynucleotides are useful in
CC diagnosing, treating and preventing diseases or conditions associated
CC with the decreased expression or overexpression of NAAP, such as cell
CC proliferative (e.g. cancer, atherosclerosis), neurological (e.g.
CC epilepsy, Huntington's disease, stroke), immune/inflammatory (e.g. AIDS,
CC allergies) and developmental (e.g. Hypothyroidism, Cushing's syndrome)
CC disorders, or infections. These are also useful in assessing the effects
CC of exogenous compounds on the expression of nucleic acid and amino acid
CC sequences of NAAP. The NAAP or its fragments are useful in screening
CC compounds for effectiveness as agonist or antagonist of the polypeptides,
CC or in altering the expression of the target polynucleotide and compounds
CC that specifically bind to or modulate the activity of the polypeptide.
CC The microarray is useful in monitoring or measuring protein-protein
CC interactions, drug-target interactions, and gene expression profiles.
CC This sequence encodes a novel human nucleic acid-associated protein
CC (NAAP)
XX
XX SQ Sequence 1705 BP; 622 A; 305 C; 314 G; 464 T; 0 U; 0 Other;

Query Match 7.8%; Score 96.8; DB 9; Length 1705;
Best Local Similarity 52.9%; Pred. No. 1.5e-09;
Matches 259; Conservative 0; Mismatches 222; Indels 9; Gaps 2;

QY 7 GAAAGTGAAGCAACCAATCGATATCATCTTTAATCTCTCTCTTCATCAGCTCC 66
DB 695 GAATGTCCCAAAACTTTTGGATGCTTTCACACCTTAACATTAAGAAATCCAAACT 754
QY 67 AAAAGTATATTTGACATATGAAAGGATGATTAAGCTTATATCGACCATATTATTA 126
DB 755 AGAGTAATTTCTCAAAATGTGAAGCATTAAGAGAGCTTTAACTGTCTCAACCTT 814

KW effector-like polypeptide; cancer; immunopathology; neuropathology;
 KW drug development; toxicology; carcinogenicity;
 KW signalling pathway polypeptide; adrenal gland; bladder; bone;
 KW bone marrow; brain; breast; cervix; tumor; immunopathology; AIDS;
 KW diabetes; pancreatitis; osteoporosis; ulcerative colitis; neuropathology;
 KW dementia; amnesia; epilepsy; Alzheimer's disease; depression.

OS Homo sapiens.

PN US2004010136-A1.

XX 15-JAN-2004.

XX 26-NOV-2002; 2002US-00305720.

XX 30-JAN-1998; 98US-00016434.

XX (INCY-) INCYTE GENOMICS INC.

PI Au-Young J, Seilhamer DJ;

DR WPI; 2004-090520/09.

PT New composition comprising polynucleotide probes, useful as array
 PT elements in a microarray for monitoring the expression of target
 PT polynucleotides or purifying a subpopulation of mRNAs, cDNA, or genomic
 PT fragments.

PS Claim 6; SEQ ID NO 1054; 73pp; English.

XX The invention relates to a composition of polynucleotide probes
 CC comprising first polynucleotide probes comprising at least a portion of a
 CC gene encoding a receptor-like polypeptide, second polynucleotide probes
 CC comprising at least a portion of a gene encoding a transducing
 CC polypeptide and third polynucleotide probes comprising at least a portion
 CC of a gene encoding an effector-like polypeptide. The probes of the
 CC composition are useful as array elements in a microarray for monitoring
 CC the expression of target polynucleotides. The microarray is useful in the
 CC diagnosis and treatment of cancer, an immunopathology or a
 CC neuropathology. It can also be used for drug discovery and development.
 CC toxicological and carcinogenicity studies, forensics or pharmacogenomics.
 CC Microarrays can also be used for monitoring the progression of diseases
 CC that may be associated with the altered expression of signalling pathway
 CC polypeptides. The composition can also be used to purify a subpopulation
 CC of mRNAs, cDNAs, or genomic fragments in a sample. The expression profile
 CC is also useful for the diagnosis and treatment of cancer, e.g. cancers of
 CC the adrenal gland, bladder, bone, bone marrow, brain, breast or cervix,
 CC an immunopathology, e.g. AIDS, diabetes, pancreatitis, osteoporosis or
 CC ulcerative colitis, or a neuropathology, e.g. dementia, amnesia,
 CC epilepsy, Alzheimer's disease or depression. This sequence represents a
 CC human polynucleotide probe of the invention. Note: The sequence data for
 CC this patent did not form part of the printed specification but was
 CC obtained in electronic format directly from USPTO at
 CC seqdata.uspto.gov/sequence.html.

XX Sequence 2320 BP; 903 A; 423 C; 412 G; 582 T; 0 U; 0 Other;

XX Query Match 7.8%; Score 96.6; DB 12; Length 2320;

XX Best Local Similarity 51.3%; Pred. No. 1.7e-09;

XX Matches 306; Conservative 0; Mismatches 279; Indels 12; Gaps 3;

QY 69 AAAGTATATTTGACATATGAAAGGTGATTAAGGCTTAATTCAGCATTTATTATAGA 128

DB 964 AGAGAAACCTCAATGTAAAGATGTGTAAGCTTTTAAACCGATCTTCAACCTTAC 1023

QY 129 GCAACATTTAAGAACCCAGATATGATGACCGATTAATATGTACAGTGAGCATTTGTA 188

DB 1024 TACCCCTAAGAAATTCATATCTGAGAGAAACCTTCAATATG-----GAAGAATGTGG 1077

QY 189 TAAAGCATTTTCAAGAAATTCATTTGAAACACATATTTGATCATTTCCGAAAAAAA 248

DB 1078 CAAGGCTTTAAGCACTCTCAACCTTACTACATTAAGATTAATTCATTACTGAGAGAA 1137

QY 249 ACCATTCATTTGTCAGTGTGTGTAAGGGGTAAATTCCTGCAGCAACCTTGAAGAGACA 308

DB 1138 ACCCTAACAATGTATAAATGTGGAAGAACCTTTAACAGTCTGCACACCTTACACACA 1197

QY 309 TGAATTCACCCATACAAAGTCAATTTAAATGTACATTTGAAAAATTCGAAGAACATTTTA 368

DB 1198 TGAAGTAATTCATCTAGAGAGAGAAACCTTACAAATGTAAAAATGTGAAAAAGCCTTAA 1257

QY 369 TAAACA---TCATCTTTAAGACATCATATATATATCTGTTCAATGAAAAAACATTAACGTG 425

DB 1258 TCATTTCTACACACCTTACTACGACATAGATATATCATCTGAGAGAAACCTTACAAATG 1317

QY 426 TAAACATGTATTAAGATTTTCACTGACCTTCAAAATTAAGCAACATTAATTAACA 485

DB 1318 TAAAGAAATGTGTTAAGCTTTTAAACACTCTTCAACCTTAAACATTAAGATTAATCA 1377

QY 486 TCATGTGATGATCTCTGCTATTCATATGATATCTGTTGTTTAAAAATTTCCAAAC 545

DB 1378 TACTGAGAGAAAGCTTACAAATCTTAAAGAAATGTGAAAAAGCTTTTAAACAAATCTCAAA 1437

QY 546 TTGATCAGTATTTACATTTTCAATATATTAACAACCTGCATCCAAACTTAATGCTTAATG 605

DB 1438 ACTTACTGAAACATTAAGAAATTCATACCTGAGAGAAACCTTA---TGAATGTGAAAAATG 1494

QY 606 TCGTAAAGGTGTGTGGGAAAAAGTTTATCTTCAATATTTAAGTCATGATGA 662

DB 1495 TGGCAAGCTTTTAAACCACTCTCAATCTTACTAGACATTAAGAAAGTCATACAGA 1551

RESULT 7

AAS91317

ID AAS91317 standard; cDNA; 4563 BP.

AC AAS91317;

DT 13-FEB-2002 (first entry)

DE DNA encoding novel human diagnostic protein #27121.

KW Human; chromosome mapping; gene mapping; gene therapy; forensic;

KW food supplement; medical imaging; diagnostic; genetic disorder; ss.

OS Homo sapiens.

PN WO200175067-A2.

PF 11-OCT-2001.

PF 30-MAR-2001; 2001WO-US008631.

PR 31-MAR-2000; 2000US-00540217.

PR 23-AUG-2000; 2000US-00649167.

PA (HYSB-) HYSBQ INC.

PI Drmanac RT, Liu C, Tang YT;

DR WPI; 2001-639362/73.

DR P-PSDB; ABG27130.

XX New isolated polynucleotide and encoded polypeptides, useful in

XX PT diagnostics, forensics, gene mapping, identification of mutations

XX PT responsible for genetic disorders or other traits and to assess

XX PT biodiversity.

CC Claim 1; SEQ ID NO 27121; 103pp; English.

CC The invention relates to isolated polynucleotide (I) and polypeptide (II)

CC sequences. (I) is useful as hybridisation probes, polymerase chain

CC reaction (PCR) primers, oligomers, and for chromosome and gene mapping,

CC and in recombinant production of (II). The polynucleotides are also used

CC in diagnostics as expressed sequence tags for identifying expressed

CC genes. (I) is useful in gene therapy techniques to restore normal

CC activity of (II) or to treat disease states involving (II). (II) is
 CC useful for generating antibodies against it, detecting or quantitating a
 CC polypeptide in tissue, as molecular weight markers and as a food
 CC supplement. (II) and its binding partners are useful in medical imaging
 CC of sites expressing (II). (I) and (II) are useful for treating disorders
 CC involving aberrant protein expression or biological activity. The
 CC polypeptides and polynucleotide sequences have applications in
 CC diagnostics, forensics, gene mapping, identification of mutations
 CC responsible for genetic disorders or other traits to assess biodiversity
 CC and to produce other types of data and products dependent on DNA and
 CC amino acid sequences. AAS64197-AAS94564 represent novel human diagnostic
 CC coding sequences of the invention. Note: The sequence data for this
 CC patent did not appear in the printed specification, but was obtained in
 CC electronic format directly from WIPO at
 CC ftp.wipo.int/pub/published_pct_sequences

80 Sequence 4563 BP; 1752 A; 814 C; 824 G; 1173 T; 0 U; 0 Other;
 Query Match 7.7%; Score 95.2; DB 5; Length 4563;
 Best Local Similarity 52.7%; Pred. No. 3.5e-09;
 Matches 258; Conservative 0; Mismatches 223; Indels 9; Gaps 2;

QY 7 GAAGTGCAGAAACCAATGATATCATCTTAAATCTTCTTCTGATCAGTCCC 66
 DB 1315 GAATGTGGCAAGCCTTTAACAGCTCCTCACTTCAACATAAACAATTCATCT 1374
 QY 67 AAAAAATATTTTGCATATATGAAGGTGTGATTAAGCCATTAATGACCATCTATTA 126
 DB 1375 GGAGGGAAAACTCAACAATGTGAAGATGTGGCAAGCTTTTAAACATCTTTAGCCCTT 1434
 QY 127 GAGCAACATTTAAGAACCCACAGTATGATCGACCGTAAATGTATGACGAGATTTG 186
 DB 1435 ACTAAACATTAAGTATATCTATCTGGGAGAAACATCAAAATGT-----GAAGATGT 1488
 QY 187 GATTAAGCATTTTTCAGAAAAATCATTTGAAAAACATATTTGATCATCTCCGAAAA 246
 DB 1489 GGGAAAGCTTTTAGGCAATCCTCAACCTTACTAGATTAAGCAATTCATCTGGAG 1548
 QY 247 AAACCATTCATTTGTCAGTGTGTGTGAAGGGTTAATCTGCAACAACCTGAAAAA 306
 DB 1549 AAACCCCTCAAAATGTGAAGATGTGGCAAGCTTTTAAACATTTCTCAGACCTTAGAGA 1608
 QY 307 CATGAATCACCCATCAAAAGTCATTTAAATGTACATTTGAAAAATGTCAAGAAGCATTT 366
 DB 1609 CATTAAGTATTTCTACTGAAAAAGAACCTTCAAAATGTGAAGATGTGGAAAAAGCTTTT 1668
 QY 367 TATTAACATCATCTTTAAGA---CATCATATATTTATCTGTGATGAAAAAAGCATTAACG 423
 DB 1669 AGCGAGTCTCAACCTTGAAGAACCATCATATATTGATCTGGAAGAAAAACCTTCAAA 1728
 QY 424 TGTAAACATGTATTAAGTTTTCATCTGACCTTCAAAATTTAGCAACAATTAATTA 483
 DB 1729 TGTGAAGAATGTGTAAAGCTTTTAAAGTGTCAATCAAAACTTGTATGATAAGTAAAT 1788
 QY 484 CATCATGCTG 493
 DB 1789 CATACTGGAG 1798

RESULT 8
 ID ADN04717 standard; cDNA; 981 BP.
 AC ADN04717;
 XX
 DT 01-JUL-2004 (first entry)
 XX
 DE Antipsoriatic cDNA sequence #572.
 XX
 KW de; gene; antipsoriatic; gene therapy; psoriasis; diagnosis.
 XX
 OS Homo sapiens.
 XX

PN WO2004028479-A2.
 XX
 PD 08-APR-2004.
 XX
 PF 25-SEP-2003; 2003WO-US030907.
 XX
 PR 25-SEP-2002; 2002US-0414006P.
 XX
 PA (GENTH) GENENTECH INC.
 XX
 PI Bodary S, Clark H, Jackman J, Schoenfeld J, Williams PM, Wood WT;
 PI Wu TD;
 XX
 DR WPI; 2004-305105/28.
 DR P-PDSB; ADN04718.
 XX
 PT New PRO nucleic acid or polypeptide, useful for preparing a
 PT pharmaceutical composition for diagnosing or treating psoriasis in a
 PT mammal.
 PS
 XX Claim 1; SEQ ID NO 1111; 3069bp; English.
 CC The invention relates to novel polynucleotide and polypeptides for
 CC treating psoriasis or a sequence having at least 80% identity to the
 CC above sequences. The nucleic acid is useful for preparing a composition
 CC for diagnosing or treating psoriasis in a mammal. This sequence
 CC corresponds to one of the polynucleotides of the invention.

80 Sequence 981 BP; 380 A; 178 C; 170 G; 253 T; 0 U; 0 Other;
 Query Match 7.7%; Score 95; DB 12; Length 981;
 Best Local Similarity 50.4%; Pred. No. 3e-09;
 Matches 289; Conservative 0; Mismatches 275; Indels 9; Gaps 2;

QY 89 AAGGCTGTGATTAAGGCTTAATATGACATCATATTATTAAGCAACATTAAGAACCACA 148
 DB 1 AAGAAATGTGGCAAGCCTTAGATATATCTCAACCCCTTCAAAACATTAAGATTAATCATA 60
 QY 149 GTATATATGACCGTATTAATGATGACGTGAGCATTTGTATTAAGCATTTTCAAGAAAT 208
 DB 61 CTGGAGAGAAACCTTCAAAATGCA-----ATGAATGTGTGAAGCCTTTAACTGTGCT 114
 QY 209 CACATTTGAAACACATATTGTATCATCTCCGAAAAAAGCATTCATCTGTCAGTGT 268
 DB 115 CAACCTTACTTAACATTAAGATTCATCTGAGAGAAACCTTCAAAATGTGAAGAAAT 174
 QY 269 GTGTAAAGGGGTTAATTTCTGCAACACCTTGAAGAAAGACATGAATCACCCATCAAAAGT 328
 DB 175 GTGGCAAGGCTTTTAAACAGTCCCTCAACCTTACTAGACATTAAGATTAAGTCACTGAG 234
 QY 329 CATTAAATGTACATTTGAAGATTTGCAAGAGCATTTTATTAACATCAATCTTTAAGAC 388
 DB 235 AGAAACCTTCAAAATGTGAAGATGTGTAAAGCCTTTAAACGATCCACAAACCTTAAGAC 294
 QY 389 ATCATATATATTTCTGTCATGAAGAAACATTAACGTTAAGCAATGAATTAATTAAGTTTCA 448
 DB 295 ATTAAGAAATTTATTAAGTAAAGAAACCAATCAAAATGTGAAGATGTGAAGAAAGCCTTTA 354
 QY 449 CTCGACCTTCAAAATTAAGCAACATTAATTAAGATCATGATGATGATCTCTGCTTATC 508
 DB 355 GTGTATTTCTCAACCTTACTTAACATTAAGTATTTACTACTGAGCAAAACCTTACAAAT 414
 QY 509 AATGTGATCATCTGTGTTGTTTAAAAATTTCCAAACCTTGTCAGTATTAACAATTTCAAT 568
 DB 415 GTGACGAATGTGGCAAGTGCCTTTAGGGCAATTCCAACCTTCAAGCAATTAAGAGAGTTC 474
 QY 569 TAAACAACGTGATCCAAAACTTAATGTCCTAATGTGTAAAGTTGTGTTGGAAAAA 628
 DB 475 ATACTGGAAGAAACCTTAC---AAATGCAATGAATGTGTAAAGCCTTTAACTGTGCT 531
 QY 629 AAGGTTATCTTCAATATGATTAAGTCATGATG 661
 DB 532 CAACCTTACTTAACATTAAGAGATTCATACCTG 564

RESULT 9
ID AAL51569
AAL51569 standard; DNA; 2597 BP.
XX
AC AAL51569;
XX
DT 10-APR-2003 (first entry)
XX
DE Human nucleic acid-associated protein coding sequence - SEQ ID NO 52.
XX
KM Human; gene; de; nucleic acid-associated protein; NAAP; arteriosclerosis;
cell proliferative disorder; atherosclerosis; cirrhosis; hepatitis; AIDS;
cancer; developmental disorder; renal tubular acidosis; anaemia; asplenia;
mental retardation; neurological disorder; Alzheimer's disease; epilepsy;
Parkinson's disease; autoimmune disorder; inflammatory disorder; allergy;
Crohn's disease; transgenic animal; animal model.
XX
OS Homo sapiens.
XX
PN WO2003000864-A2.
XX
PD 03-JAN-2003.
XX
PF 20-JUN-2002; 2002WO-US021179.
XX
PR 22-JUN-2001; 2001US-0300518P.
PR 29-JUN-2001; 2001US-0301787P.
PR 29-JUN-2001; 2001US-0301792P.
PR 29-JUN-2001; 2001US-0301892P.
PR 29-JUN-2001; 2001US-0301933P.
PR 06-JUL-2001; 2001US-0303405P.
PR 06-JUL-2001; 2001US-0303442P.
PR 15-MAR-2002; 2002US-0364438P.
XX
PI (INCY-) INCYTE GENOMICS INC.
XX
PI Gandhi AR, Swarnakar A, Hafalia AJA, Warren BA, Emerling BM;
PI Arvizu CS, Ison CH, Honchell CD, Lee EA, Yue H, Forsythe IU;
PI Ramkumar J, Griffin JA, Yang J, Sanjanwala NM, Baughn MR;
PI Borowsky ML, Yao MG, Wallia NK, Bandman O, Lai PG, Becha SD, Lee SY;
PI Richardson TW, Elliott VS, Luo W, Tang YT, Zebardjian Y, Lu Y;
XX
DR MPI; 2003-201420/19.
DR P-PSDB; AAO16419.
XX
PT New nucleic acid-associated proteins and polynucleotides, useful for
diagnosing, treating or preventing cell proliferative (e.g. cancer),
PT neurological (e.g. epilepsy or Parkinson's disease), or autoimmune
PT disorders (e.g. AIDS).
XX
PS Claim 12; Page 292-293; 312pp; English.
XX
XX The invention comprises the amino acid and coding sequences of human
nucleic acid-associated proteins (NAAP). The DNA and protein sequences of
the invention are useful for diagnosing, treating or preventing disorders
associated with aberrant expression of NAAP, such as: cell proliferative
disorders (e.g. arteriosclerosis, atherosclerosis, cirrhosis, hepatitis
or cancer); developmental disorders (e.g. renal tubular acidosis, anaemia
or mental retardation); neurological disorders (e.g. Alzheimer's disease,
Parkinson's disease or epilepsy); and autoimmune/inflammatory disorders
(e.g. AIDS, allergies, asthma or Crohn's disease). The DNA sequences of
the invention are useful for creating transgenic animals to model human
disease. The present DNA sequence encodes a human nucleic acid-associated
protein of the invention
XX
SQ Sequence 2597 BP; 947 A; 504 C; 505 G; 641 T; 0 U; 0 Other;
XX

QY 78 TTGCACATATGAGAGGGTGTGATTAAGCCTATATATCAGCCATCATTTATAGCAACATTT 137
Db 1070 TTATATAATGTGAAGATGTGGCAAAAGCCTTATATGTCTCAACCTTACTCAACATTA 1129
QY 138 AAGAACCCACGATATGATTCGACCGGTATTAATGTACAGTGGACGATTTGGATTAAGCAT 197
Db 1130 GAGAAATTCATCTGAGAGAAACCTTACAAATGT-----GAAAGTGTGGCAAAAGCCTT 1183
QY 198 TTTTCAGAAAATTCATCTTGGAAAACACATATTTGTATCATATTTCCGAAAAAAACCATTTCCA 257
Db 1164 TTAAGTGTCTTCACTTCTTCACTCAACATTAAGAAATTTCTACTGTGAGAAAAACCTATGAA 1243
QY 258 TTGTCAGTGTGTGTGTAAGGGGTTAATTTCTGACACACATTTGAAAGACATGAATGAC 317
Db 1244 ATGTAAAGAAATGTGGCAAAAGCCTTAAACATCTTCAACCTGACACACATTAAGAAAT 1303
QY 318 CCATCAAAAGTATTAATTAATGTACATTGTAATTTGTCAAGAAAGATTTTAATA--ACA 374
Db 1304 TCATACCGGAGAGAAACCTTACAAATGTGAAGAAATGTGGCAAAAGCCTTAAACCAATTTCTC 1363
QY 375 TCAATCTTTAAGACATCATATTAATTAATCTGTTTCATGAAAAAATTAAGGTGAACAATG 434
Db 1364 ACACTTACTACATCACTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1423
QY 435 TATTAAGTTTCTACTGACCTTCAAAATTAAGACACATTAATTAATTAATTAATTAATTA 493
Db 1424 TGGCAAAAGCTTTAAGCGGTCTCTCAAACTTACTGAACATATGATTAATTAATTAATTA 1482
RESULT 10
AAS64586
ID AAS64586 standard; cDNA; 3639 BP.
XX
AC AAS64586;
XX
DT 13-FEB-2002 (first entry)
XX
DE DNA encoding novel human diagnostic protein #390.
XX
KM Human; chromosome mapping; gene mapping; gene therapy; forensic;
KM food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX
OS Homo sapiens.
XX
PN WO200175067-A2.
XX
PD 11-OCT-2001.
XX
PF 30-MAR-2001; 2001WO-US008631.
XX
PR 31-MAR-2000; 2000US-00540217.
PR 23-AUG-2000; 2000US-00649167.
XX
FA (HISE-) HISEQ INC.
XX
PI Drmanac RT, Liu C, Tang YT;
XX
DR MPI; 2001-639362/73.
DR P-PSDB; ABB00399.
XX
PT New isolated polynucleotide and encoded polypeptides, useful in
diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity.
XX
XX Claim 1; SEQ ID NO 390; 103bp; English.
XX
XX The invention relates to isolated polynucleotide (I) and polypeptide (II)
sequences. (I) is useful as hybridisation probes, polymerase chain
reaction (PCR) primers, oligomers, and for chromosome and gene mapping,
CC and in recombinant production of (II). The polynucleotides are also used
in diagnostics as expressed sequence tags for identifying expressed
genes. (I) is useful in gene therapy techniques to restore normal

CC activity of (II) or to treat disease states involving (II). (II) is
 CC useful for generating antibodies against it, detecting or quantifying a
 CC polypeptide in tissue, as molecular weight markers and as a food
 CC supplement. (II) and its binding partners are useful in medical imaging
 CC of sites expressing (II). (I) and (II) are useful for treating disorders
 CC involving aberrant protein expression or biological activity. The
 CC polypeptides and polynucleotide sequences have applications in
 CC diagnostics, forensics, gene mapping, identification of mutations
 CC responsible for genetic disorders or other traits to assess biodiversity
 CC and to produce other types of data and products dependent on DNA and
 CC amino acid sequences. AAS64197-AAS94564 represent novel human diagnostic
 CC coding sequences of the invention. Note: The sequence data for this
 CC patent did not appear in the printed specification, but was obtained in
 CC electronic format directly from WIPO at
 CC ftp.wipo.int/pub/published_pct_sequences
 XX
 SQ Sequence 3639 BP; 1403 A; 708 C; 622 G; 906 T; 0 U; 0 Other;
 Query Match 7.5%; Score 92.4; DB 5; Length 3639;
 Best Local Similarity 54.1%; Pred. No. 1.2e-08;
 Matches 236; Conservative 0; Mismatches 191; Indels 9; Gaps 2;
 QY 69 AAGATATATTTGCACATATAGAGGTGTGATTAAGCTTATATGACCATCATTTATTAGA 128
 DB 738 AGAGAAACCTTACAAATGTGAAGAAATGTGGCAAGCTTTAGCATCTTCAACCTTGC 797
 QY 129 GCACATTTAAGAACCCACAGTATGATTCACCGTATTAATGTAAGTGAAGCATTTGTGA 188
 DB 798 TAAACATTAAGAAATTTATCTGAGAGAAACCTTAAATGT-----GAAAGATGTGG 851
 QY 189 TAAAGATTTTTCGAAATATCATTTGGAACACATATTGTATCATATTCGAAATAA 248
 DB 852 CAAAGCTTTTACCGCTTCTTCAACCTTGTCAAAATTAAGAAATTTACTGTGAGAGAA 911
 QY 249 ACCATTCATTTGTGAGTGTGTGTAAGAGGGTTATTTCTGACACACCTTGAAGAGAA 308
 DB 912 ACCCTACAAATGTAAAGAAATGTGGCAAGCTTTAGCAATTCCTCAACCTTGTCTATCA 971
 QY 309 TGAATATACCCATTAAGATCATTTAAATGTACATTTGAAATTTGTCAAGAGATTTTA 368
 DB 972 TAAAGATACCTATATCTGAAGAAACCTTCAATATGTAAGAAATGTCAAAAGCTTTTAA 1031
 QY 369 TAAACATCATTC--TTTAAGACATCATATATTTATCTGTCATGAATAAAACATTAAGCTG 425
 DB 1032 GCGACTCTCAACCTTACTTAACATTAATTAATATATGCTGAGAGAAACCTTCAATG 1091
 QY 426 TAAACATGTATTAAGTTTCACTCGACCTTCAAAATTAGCAACATTAATTTAAACA 485
 DB 1092 TGAAGAAATGTGGCAAGCTTTTATGATTCATCTTCAAACTTACTATATAGTTTATTC 1151
 QY 486 TCATGTGTGATTCCT 501
 DB 1152 TACTGGAGAGAAACCT 1167
 RESULT 11
 ACH87568/C
 ID ACH87568 standard; DNA; 994 BP.
 XX
 AC ACH87568;
 XX
 DT 29-JUL-2004 (first entry)
 XX
 DE Human genome derived single exon probe #20763.
 XX
 KM Human; probe; 86; gene expression; single exon probe; microarray;
 XX
 KM alternative splicing event; genomic alteration.
 OS Homo sapiens.
 XX
 XX US2003194704-A1.
 XX
 PD 16-OCT-2003.

XX
 PF 03-APR-2002; 2002US-00029386.
 XX
 PR 03-APR-2002; 2002US-00029386.
 XX
 PA (PENN/) PENN S G.
 PA (RANK/) RANK D R.
 PA (HANZ/) HANZEL D K.
 XX
 PI Penn SG, Rank DR, Hanzel DK;
 XX
 DR WPI; 2004-119264/12.
 XX
 PT New human genome-derived single exon nucleic acid probes useful for human
 PT gene expression analysis, for identifying or characterizing alternative
 PT splicing events, for assessing genomic alterations or as tools for
 PT surveying tissues.
 XX
 PS Claim 1; SEQ ID NO 20763; 80pp; English.
 XX
 CC The invention relates to a nucleic acid probe for measuring human gene
 CC expression, comprising any of the 27,400 fully defined nucleotide
 CC sequences in the specification, or their complements or fragments, and
 CC encoding at least 8 amino acids of any of the 6888 amino acid sequences
 CC fully defined in the specification. The probe is a single exon probe that
 CC hybridizes under high stringency conditions to a nucleic acid molecule
 CC expressed in human cells or tissues. Also included are a spatially-
 CC addressable set of single exon nucleic acid probes for measuring human
 CC gene expression (comprising a plurality of single exon nucleic acid
 CC probes cited above, where each of the plurality of probes is separately
 CC and addressably isolatable or amplifiable from the plurality), a single
 CC exon microarray for measuring human gene expression, a method of
 CC measuring human gene expression, a vector comprising the single exon
 CC probe cited above, an ORF-encoded peptide comprising at least 8
 CC contiguous amino acids of any of the above-mentioned amino acid
 CC sequences (optionally with conservative amino acid substitutions), an
 CC isolated antibody that binds specifically to a peptide cited above,
 CC methods of selling and/or licensing single exon probes or microarrays to
 CC a customer desiring to measure gene expression, a method of providing
 CC human gene expression data by subscription, and a computer-readable
 CC storage medium which contains a database having a plurality of records
 CC (each record including data on the expression of a single exon probe
 CC cited above. The probe, methods and apparatus are useful in gene
 CC expression analysis. The probes may be used as tools for surveying
 CC tissues to detect the presence of expressed messages that contain their
 CC specific exon, or in constructing genome-derived single exon microarrays.
 CC In addition, the probes are used in identifying and characterizing
 CC alternative splicing events, in detecting and characterizing gross
 CC alterations in the genomic locus that includes their exon, in assessing
 CC smaller genomic alterations, in priming the synthesis of nucleic acids,
 CC or in expressing the ORF-encoded peptide. The present sequence is a human
 CC single exon probe of the invention. Note: The sequence data for this
 CC patent did not form part of the printed specification, but was obtained
 CC in electronic format directly from USPTO at
 CC seqdata.uspto.gov/sequence.html?docid=2003194704
 XX
 SQ Sequence 994 BP; 252 A; 159 C; 203 G; 380 T; 0 U; 0 Other;
 Query Match 7.4%; Score 92; DB 12; Length 994;
 Best Local Similarity 52.0%; Pred. No. 1.2e-08;
 Matches 259; Conservative 0; Mismatches 230; Indels 9; Gaps 2;
 QY 7 GAAAGTGAGAAACCAATGATATCATCTTTATATCTTCTTCTTATCATCGTCC 66
 DB 763 GAATGTGGCAAGCTTCTTCAACATCTTCTTCAACCTTCAACATTAAGTATTCATCT 704
 QY 67 AAAAGTATTTTGCACATATAGAGGTGTGATTAAGCCATTAATGACCATCATTTATTA 126
 DB 703 GAGAGAAACCTTCAATATGTGAAGATGTGTAAGCTTTTAAACACCTTCAAGCCCTT 644
 QY 127 GAGCAATTTAAGAACCCACAGTATGATCGACCGTATTAATGTACAGTGAAGATGT 186
 DB 643 ACTACACATTAAGTTCAATTGATTTAAAGAAACCTTCAAAATGT-----GAAAGATGT 590

QY 187 GATTAAGCATTTTTCAGAAAATCATTGGAACATATTTGATTCATTCGAAA 246
PT 589 GACAAAGCTTTTAAACGATTCATCTTACTTAAACATTAATTCATTCGAGAG 530
Db 247 AAACCATTCATTTGTCAGTGTGGTAAAGGGGTTAATTCCTGACAAACCTGAAA 306
QY 529 AAATCTTAACAATGTGAACATGTGGCAAAAGGCTTTAACTGCTTCAACCTTA 470
Db 307 CATGAATCAACCCATCAAAAGTCATTTAAATGTACATTTGAAAATTTGCAAGAGCAT 366
QY 469 CATAAAGAATTCATCTGGAGAAAACCTTAAATGTGAAGAAATGTGGCAAAAGCTTT 410
Db 367 TATTAACAT---CAATCTTAAGACATCATATTTATCTGTTGATGAAAAAACATTA 423
QY 409 AATGTCTCTTACACCTTCTACACATTAAGATGATTCATTAAGTGAAGAAACCTTACAA 350
QY 424 TGTAAACATGTAAATTAAGTTTCACTGACCTTCAAAATTTAGCACACATTAATTA 483
Db 349 TGTGAAGAAATGTGGCAAGGCTTTAAACCACTCTCAAAACCTTAATTAATTAAT 290
QY 484 CATCATGTGTGATCTCCT 501
Db 289 CATACTGGAGAGAAACCT 272

RESULT 12
ADRI4368
ID ADRI4368 standard; DNA; 2873 BP.
AC ADRI4368;
XX
DT 21-OCT-2004 (first entry)
XX
DE Human NF-kappaB pathway-associated gene SeqIDJ369.
XX
KW NF-kappaB pathway; antiinflammatory; cytosolic; hepatotropic; virucide;
KW antitubercitic; antineuritic; gastrointestinal-gen; antiaesthetic;
KW antitubercitic; antineuritic; immunomodulator; cerebroprotective; vasotropic;
KW immunosuppressive; vulnary; gene therapy; immune disorder;
KW inflammatory disorder; NF-kappaB regulation; cancer; aberrant apoptosis;
KW hepatic disorder; Hodgkin's lymphoma; haematopoietic tumour;
KW hyper-igm syndrome; hypohidrotic ectodermal dysplasia;
KW X-linked anhidrotic ectodermal dysplasia; immunodeficiency;
KW viral infection; HIV-1; hepatitis B; hepatitis C; EBV; influenza;
KW viral replication; host cell survival; evasion of immune response;
KW rheumatoid arthritis; inflammatory bowel disease; colitis; asthma;
KW atherosclerosis; cachexia; euthyroid sick syndrome; stroke; EAE;
KW autoimmune disorder; hyper immune activity;
KW aberrant acute phase response; hypercongenital condition; birth defect;
KW necrotic lesion; wound; organ transplant rejection;
KW aberrant signal transduction; proliferating disorder; cancer;
KW HIV propagation; gene; ds; human.
XX
OS Homo sapiens.
XX
PN WO2004065577-A2.
XX
PD 05-AUG-2004.
XX
PF 13-JAN-2004; 2004MO-US000798.
XX
PR 14-JAN-2003; 2003US-0440068P.
XX
PR 12-MAY-2003; 2003US-0469757P.
XX
PA (BRIM) BRISTOL-MYERS SQUIBB CO.
XX
PI Nadler SG, Neubauer MG, Feder JN, Carman J;
XX
XX WPI; 2004-562168/54.
XX
XX P-PSDB; ADRI4369.
XX
PT New isolated polynucleotides and polypeptides associated with NF-kappaB

PT pathway, useful for diagnosing, treating, or preventing disorders or
PT diseases associated with NF-kappaB pathway.
XX
XX Claim 1, SEQ ID NO 369; 237bp; English.

CC This invention relates to the novel association of protein sequences (and
CC the genes which encode them) to the NF-kappaB pathway. The invention may
CC be useful for the production of compounds with an antiinflammatory,
CC gastrocitic, hepatotropic, virucide, antitubercitic, antineuritic,
CC gastrointestinal-gen, antiaesthetic, antiatherosclerotic,
CC immunomodulator, cerebroprotective, vasotropic, immunosuppressive or
CC vulnary activity or for gene therapy. The proteins and nucleotides are
CC useful for diagnosing, preventing, treating, or ameliorating conditions
CC or diseases associated with the NF-kappaB pathway. The condition is an
CC immune disorder, an inflammatory disorder, an inflammatory disorder
CC related to aberrant NF-kappaB regulation, cancer, aberrant apoptosis,
CC hepatic disorder, Hodgkin's lymphoma, haematopoietic tumours, hyper-igm
CC syndromes, hypohidrotic ectodermal dysplasia, X-linked anhidrotic
CC ectodermal dysplasia, immunodeficiency, viral infections, HIV-1, HTLV-1,
CC hepatitis B, hepatitis C, EBV, influenza, viral replication, host cell
CC survival, evasion of immune responses, rheumatoid arthritis, inflammatory
CC bowel disease, colitis, asthma, atherosclerosis, cachexia, euthyroid sick
CC syndrome, stroke, EAE, autoimmune disorders, disorders related to hyper
CC immune activity, disorders related to aberrant acute phase responses,
CC hypercongenital conditions, birth defects, necrotic lesions, wounds,
CC organ transplant rejection, conditions related to organ transplant
CC rejection, disorders related to aberrant signal transduction,
CC proliferating disorders, cancers and HIV propagation in cells infected
CC with other viruses. The present sequence is that of a human gene which is
CC subject to the novel association with the NF-kappaB pathway of the
CC invention. Note: This sequence does not appear in the specification but
CC was obtained by the indexer from Genbank.
XX
SQ Sequence 2873 BP; 948 A; 564 C; 545 G; 816 T; 0 U; 0 Other;

Query Match 7 4%; Score 92; DB 13; Length 2873;
Best Local Similarity 52.2%; Pred. No. 1.4e-08;
Matches 256; Conservative 0; Mismatches 225; Indels 9; Gaps 2;

QY 7 GAAATGACGAAACCAATTCATATCATTTAATATCTTCTTTCATCAGTCC 66
Db 1621 GAATGTCCGAAACATTTTGCATGCTTTCACCTTACTCAACATTAAGATCCAACT 1680
QY 67 AAAAGTATTTTGCATATGAAAGGTTGATTAAGGCTTATATGACCATATTATTA 126
Db 1681 AAGAGGAATTTCTCAAAATGTGAAGCATATGGAAGGCTTTAACTGGTCCCAACCTT 1740
QY 127 GAGCAACATTTAAGAACCAAGTATGATGACCGTATTAATGTACGTGACGATGT 186
Db 1741 AATTAACATTAAGAAATTCATCTGAGAAAACCTTCAATGTAA-----GAATGT 1794
QY 187 GATTAAGCATTTTTCAGAAAATCATTGGAACATATTTGATTCATTCGAAA 246
Db 1795 GGCAAAGCTTTTAAACGACCTCACAACCTTATTAACATTAAGAAATTCATCTGAAG 1854
QY 247 AAACCATTCATTTGTCAGTGTGGTAAAGGGTAAATTCCTGACAAACCTTGAAGA 306
Db 1855 AAACCTTAACAATGTGAAGATGTGGCAAAAGGCTTTAAACGATTCAGCCCTTACTCA 1914
QY 307 CATGAATCAACCCATCAAAAGTCATTTAAATGTACATTTGAAAATTTGCAAGAGCAT 366
Db 1915 CATTAATTAATTCATATCTGGGAAATTCCTTCAAAATGTGAAGAAATGTGTAAGACCTTT 1974
QY 367 TATTAACATCATCTTTAAGACATCATATTAATATCTGT---CATGAAGAAACATTAAG 423
Db 1975 AACCAAGCTCAAGACCTTACTGAAACATTAAGTTAATTCATCCGGAAGAAAGTTATGA 2034
QY 424 TGTAAACATGTAAATTAAGTTTCACTGACCTTCAAAATTTAGCAACATTAATTA 483
Db 2035 TGTGAAGATGTGGCAAGGCTTTTAAACGATCTCAAAACCTTACTGAACATTAAGATTA 2094
QY 484 CATCATGTGT 493

Db 2095 CATACTGAG 2104

RESULT 13

ID ADN99064 standard; cDNA; 1533 BP.

XX ADN99064;

XX AC

XX ADN99064;

XX DT 29-JUL-2004 (first entry)

XX DE Novel human cDNA sequence #664.

XX de; gene; anti-inflammatory; dermatological; neuroprotective;

KM immunomodulator; antibacterial; virucide; antiproliferic; cytostatic;

KM gene therapy; vaccine; inflammatory; CNS; immune disorder; cancer;

KM psoriasis; diabetes; early aging; hormonal imbalance;

KM ischemic heart disease; ulcerative colitis.

XX Homo sapiens.

XX PN WO2004038003-A2.

XX PD 06-MAY-2004.

XX PF 24-OCT-2003; 2003MO-US033947.

XX PR 25-OCT-2002; 2002US-0421061P.

PR 25-OCT-2002; 2002US-0421080P.

PR 25-OCT-2002; 2002US-0421552P.

PR 25-OCT-2002; 2002US-0421614P.

PR 30-OCT-2002; 2002US-0422177P.

PR 30-OCT-2002; 2002US-0422178P.

PR 15-NOV-2002; 2002US-0426355P.

PR 15-NOV-2002; 2002US-0426384P.

PR 15-NOV-2002; 2002US-0426394P.

PR 15-NOV-2002; 2002US-0426430P.

PR 15-NOV-2002; 2002US-0426916P.

PR 27-NOV-2002; 2002US-0429224P.

PR 27-NOV-2002; 2002US-0429275P.

PR 27-NOV-2002; 2002US-0429302P.

PR 27-NOV-2002; 2002US-0429366P.

PR 27-NOV-2002; 2002US-0429651P.

PR 04-DEC-2002; 2002US-0430645P.

PR 04-DEC-2002; 2002US-0430651P.

PR 04-DEC-2002; 2002US-0430657P.

PR 04-DEC-2002; 2002US-0430663P.

PR 04-DEC-2002; 2002US-0430684P.

PR 04-DEC-2002; 2002US-0430688P.

PR 05-DEC-2002; 2002US-0430937P.

PR 05-DEC-2002; 2002US-0430965P.

PR 05-DEC-2002; 2002US-0431458P.

PR 12-DEC-2002; 2002US-0433251P.

PR 12-DEC-2002; 2002US-0433500P.

PR 13-DEC-2002; 2002US-0433316P.

PR 13-DEC-2002; 2002US-0433318P.

PR 23-DEC-2002; 2002US-0436236P.

PR 03-JAN-2003; 2003US-0437914P.

PR 17-JAN-2003; 2003US-0440820P.

PR 17-JAN-2003; 2003US-0440821P.

PR 18-APR-2003; 2003US-0463700P.

PR 18-APR-2003; 2003US-0463708P.

PR 18-APR-2003; 2003US-0463716P.

PR 18-APR-2003; 2003US-0463732P.

PR 02-MAY-2003; 2003US-0467199P.

PR 02-MAY-2003; 2003US-0467201P.

PR 02-MAY-2003; 2003US-0467203P.

PR 02-MAY-2003; 2003US-0467230P.

PR 19-MAY-2003; 2003US-0471306P.

PR 19-MAY-2003; 2003US-0471336P.

PR 22-MAY-2003; 2003US-0472420P.

PR 22-MAY-2003; 2003US-0472430P.

PR 09-JUN-2003; 2003US-0476609P.

PR 09-JUN-2003; 2003US-0476621P.

PR 09-JUN-2003; 2003US-0476632P.

PR 09-JUN-2003; 2003US-0476641P.

PR 08-JUL-2003; 2003US-0485217P.

PR 08-JUL-2003; 2003US-0485218P.

PR 08-JUL-2003; 2003US-0485218P.

PR 08-JUL-2003; 2003US-0485223P.

PR 08-JUL-2003; 2003US-0485224P.

PR 08-JUL-2003; 2003US-0485252P.

PR 08-JUL-2003; 2003US-0485259P.

PR 14-JUL-2003; 2003US-0486446P.

PR 14-JUL-2003; 2003US-0486480P.

PR 15-JUL-2003; 2003US-0486891P.

PR 15-JUL-2003; 2003US-0486960P.

PR 08-AUG-2003; 2003US-0493341P.

PR 08-AUG-2003; 2003US-0493370P.

PR 08-AUG-2003; 2003US-0493573P.

PR 08-AUG-2003; 2003US-0493577P.

XX (FIVE-) FIVE PRIME THERAPEUTICS INC.

XX PA

XX PI Williams LT, Chu K, Lee E, Hestir K, Beaurang PA, Behrens D;

PI Halendeck RF, Kochakota S, Lin H, Linnemann T, Pierce K, Wang Y;

PI Wong JGP, Wu G, Zhang H, Zeng C;

XX DR WPI; 2004-36511/34.

DR P-PSDB; ADN9848.

XX PT New nucleic acid molecules, useful in preparing a composition for

PT treating or preventing e.g. inflammatory, CNS, bacterial or viral

PT disorders, cancer, psoriasis, diabetes, ischemic heart disease or

PT ulcerative colitis.

XX PS Claim 1; SEQ ID NO 664; 532pp; English.

XX CC The invention relates to a nucleic acid molecule comprising a

CC polynucleotide sequence or its complement that encodes a polypeptide. The

CC nucleic acid is useful in preparing a composition for treating or

CC preventing inflammatory, CNS, immune, bacterial or viral disorder,

CC cancer, psoriasis, diabetes, early aging, hormonal imbalance, ischemic

CC heart disease or ulcerative colitis. This sequence corresponds to a

XX nucleic acid of the invention.

XX SQ Sequence 1533 BP; 614 A; 280 C; 251 G; 388 T; 0 U; 0 Other;

Query Match 7.4%; Score 91.8; DB 12; Length 1533;

Best Local Similarity 52.1%; Pred. No. 1.4e-08;

Matches 257; Conservative 0; Mismatches 227; Indels 9; Gaps 2;

QY 11 GTGACGAACCAATGATATATCTTTAATCTCTTCTTATCATACGTCACAAA 70

Db 305 GTGGCAATCATATTTGATGCTTTTCAATTTAATCAACATAAAGAAATTCATCTAGAG 364

QY 71 AGTATTTTGCACATATGAAGGATGATTAAGCTTATATGACCATATTTAGAGC 130

Db 365 AGTATCTTACAAATGTGAAGATGTGAAGCTTTAACTGTCTCAACCTTAACTA 424

QY 131 AACATTAAAGAACCCACAGTATGATGACCGTATTAATGATACGTGACAGATTGATA 190

Db 425 AACATAAGATTAATTCATCTACTGAGAAAAACCTTACAAATGT-----GAAGATGTGGCA 478

QY 191 AAGCATTTTTCGAANAATCATCTTGGAAACATATTTGATGATCATTCGGAATAAAAC 250

Db 479 AAGCTTTTAAACGGCTCTCAAAATCTTAAATTAATAAATTAATTCATCTGAGAGAAAC 538

QY 251 CATTCATTTGATGATGATGATGAAGGATTAATCTGACAAACCTGAAAGACATG 310

Db 539 CTTAGAAATGTGAAGATGTGGAAAGCTTTTAAACGGCTTCAACCTTAACTAATCA 598

QY 311 AATCAACCATTAAGATCAATTTAAATGATGATGATGATGATGATGATGATGATGAT 370

Db 599 AAAGATTCATACAGAAAGAAACCTTACAAATGTGAAGATGTGCAAGGCTTTAAC 658

QY 371 AACATCAATCTT---AAGATCATATATTTATCTGTTCATGAAAAACATTAACTGTA 427

QY 251 CATTCATTGTCAGTGTGTAAAGGGGTAAATTCGACACACCTTGAAAAGACATG 310
DB 539 CTTACAAATGTGAAGATGTGGCAAGCTTTAAACGGGTCTCAACCTTACTAAACATA 598
QY 311 AATTCACCATTAACAAAGTCAATTTAAATGATCATTGAAAATGTCAAAGACATTTTAA 370
DB 599 AAGAAATTCATACAGAAAGAAACCCCTACAAATGTGAAGATGTGCAAGGCTTTAAAC 658
QY 371 AACATCAATCTT---AAGACATCATATATATCTGTTCATGAAAAACATTAAAGTGA 427
DB 659 AGTTCGATTTCTTAATTAACATTAAGAAATTCATATGAAATTAACCTTACAAATGTG 718
QY 428 AACATGATATAAAGTTTCTGACCTGACCTTCAAAATTAGCAACAATTAATTAAACATC 487
DB 719 AAGAAATGTGCAAAAGCTTTAGAGTATTCCTCAATTTCTTAAAAACATTAAGATTAATCATA 778
QY 488 ATGTGTGATCTCC 500
DB 779 CTGGGAAAAAAC 791

RESULT 15
ADB63624
ID ADB63624 standard; cDNA; 2110 BP.
XX
AC ADB63624;
XX
DT 04-DEC-2003 (first entry)
XX
DE Human cDNA encoding clone THYM20071120.
XX
KM Human; ss; gene; pharmaceutical; diagnostic; gene therapy;
KM tissue regeneration; cell regeneration; membrane protein;
KM signal transduction-related protein; transcription-related protein;
KM osteoporosis; neurological disease; cancer; tumour.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT CDS 170..1792
FT /tag= a
FT /product= "Clone THYM20071120 protein"

XX BP1308459-A2.
XX
XX PD 07-MAY-2003.
XX PF 28-MAR-2002; 2002BP-00007401.
XX PR 05-NOV-2001; 2001JP-00379298.
XX PR 25-JAN-2002; 2002US-00350978.
XX
XX PA (HELI-) HELIX RES INST.
XX PA (REAS-) RES ASSOC BIOTECHNOLOGY.
XX
PI Iisogai T, Sugiyama T, Otsuki T, Makamatsu A, Sato H, Ishii S,
PI Yamamoto J, Isono Y, Hito Y, Otsuka K, Nagai K, Irie R, Tamechika I,
PI Seki N, Yoshikawa T, Otsuka M, Negahari K, Masuno Y;
XX
XX WPI; 2003-450961/43.
XX DR P-PSDB; ADB65594.
XX
XX PT New polynucleotides and polypeptides, useful for developing a diagnostic
XX PT marker or medicines for regulation of their expression and activity, or
XX PT as targets of gene therapy.
XX
XX PS Claim 1; Page; 222pp; English.
XX
XX The invention discloses a polynucleotide comprising a sequence selected
XX CC from 1970 fully defined nucleotide sequences which encode novel
XX CC polypeptides. Also claimed is a polypeptide encoded by the polynucleotide
XX CC or its partial peptide, an antibody binding to the polypeptide or peptide
XX CC of the polynucleotide, immunologically assaying the polypeptide or

CC peptide of the polynucleotide by contacting the polypeptide or peptide
CC with the antibody of the encoded protein, and observing the binding
CC between the two, a transformant carrying the polynucleotide in an
CC expressible manner and an antisense polynucleotide. The oligonucleotide
CC is useful as a primer for synthesizing the polynucleotide, or as a probe
CC for detecting the polynucleotide. The polynucleotides and encoded
CC proteins are useful as pharmaceutical agents and many disease-related
CC genes may be included in them, for developing a diagnostic marker or
CC medicines for regulation of their expression and activity, or as targets
CC of gene therapy. The genes are involved in tissue and/or cell
CC regeneration. Membrane proteins, signal transduction-related proteins,
CC transcription-related proteins, disease-related proteins and genes
CC encoding them can be used as indicators for diseases (e.g. osteoporosis,
CC neurological diseases, cancer, tumour). The cDNA may be used to regulate
CC the activity or expression of the encoded protein to treat diseases. The
CC sequence presented is a cDNA of the invention. Note: Some of the sequence
CC data for this patent is not represented in the printed specification, but
CC is based on sequence information supplied by the European Patent Office.
XX
SQ Sequence 2110 BP; 784 A; 394 C; 386 G; 546 T; 0 U; 0 Other;

Query Match 7.4%; Score 91.8; DB 10; Length 2110;
Best Local Similarity 51.7%; Pred. No. 1.5e-08;
Matches 314; Conservative 0; Mismatches 277; Indels 16; Gaps 4;

QY 61 CGTCCCAAAAGATATTGTCATATGAAAGGCTGATTAAGCCATATGACCATCA 120
DB 1253 CATCCGAAGAGAAACCTTCAAAATGTAAGATGTGGCAAGCTTTAAACCTCTCA 1312
QY 121 TTATTAGCAACATTTAAGAAACCCACAGTAATGATGACCGTATTAATGTACAGTGAC 180
DB 1313 GCCCTTACATACATTAAGAAATTCACACTGAGAGAAACCCACAAATGT-----GAA 1366
QY 181 GATTGTGATTAAGACATTTTTCAGAAATCAATTTGAAACACATATTTGATACATTC 240
DB 1367 GAATGTGCAAAAGCTTTAACCAGATCTCAAAACCTTACGAACATTAAGAAACCTTACTACT 1426
QY 241 GAAAAAAACCAATTCATGTTTCAGTGTGTAAGGGGTTAATCTGCAACAACCTTG 300
DB 1427 GGAAGAAACCCCTCAAAATGTGAAGATGTGGCAAGCTTTTATCCAAATCTTCAAACTT 1486
QY 301 AAAAGACATGAATCAACCCATCAAAAGTCAATTTAATGATCAATTTGAAATTTGTCAAGAA 360
DB 1487 ACTGAACATTAATAAATTTCTTGTGAGAGATACCTTACAAAGTGAAGATGTGGCAAA 1546
QY 361 GCATTTATTAACATCAATCTTTA---AGACATCATATATATATCTGTCATGAAAAACA 417
DB 1547 GCTTTTAAACATCTCATCTCTTACTACACATTAAGAAATTCATCTGGGAGAAACCC 1606
QY 418 TTAAGCTGTAACAAATGATTAAGTTTTCATCCGACCTTCAAAATTTAGCAACAACATA 477
DB 1607 TACAAATGTGAAGATGTGGCAAGCTTTTACCGATCTTCAAACTTACTGAACATTAAG 1666
QY 478 TTAACATCATGAGTG-GATCTCTGCTTATCAATGATGATCATCTGTTGTTTAAAAA 536
DB 1667 ATTAATTCATACCTGAGAGAAACCTTAAATGTGAAGATGTGACAAAGCTTTTAAACAA 1726
QY 537 TTTCCAACTGTGTGATTAATCAATTTCAATATTAACAACTGATCAAAACTTTAATG 596
DB 1727 TCTGCAAACT-----TACTAAACATTAATAAATAATCATCTGAGAGAAATACAGAAC 1780
QY 597 TCTTAATGTGTGAAGGTTGTGTGGAAAAAGTTTATCTTCAATATGTTAAGTCA 656
DB 1781 TGGAAATGTGATTAAGATTTTGAACAACCTTCAAAATTTTCTTAATTAAGAAATCA 1840
QY 657 TGATGAT 663
DB 1841 TACTGAT 1847

Search completed: February 9, 2005, 02:39:51
Job time : 644.278 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Comugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 9, 2005, 01:51:00 ; Search time 224.272 Seconds
(without alignments)
9039.698 Million cell updates/sec

Title: US-09-831-804-2
Perfect score: 1239
Sequence: 1 atgagtgaggaagcagcgaac.....catcagtgatcttcgataa 1239

Scoring table: IDENTITY_NUC
Gapop 10.0, Gapext 1.0

Searched: 1202784 seqs, 818138359 residues

Total number of hits satisfying chosen parameters: 2405568

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents NA:
1: /cgn2_6/prodata/1/ina/5A_COMB.seq:*
2: /cgn2_6/prodata/1/ina/5B_COMB.seq:*
3: /cgn2_6/prodata/1/ina/6A_COMB.seq:*
4: /cgn2_6/prodata/1/ina/6B_COMB.seq:*
5: /cgn2_6/prodata/1/ina/PCTUS_COMB.seq:*
6: /cgn2_6/prodata/1/ina/backfillseq.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	584.8	47.2	627	4 US-09-248-796A-5184	Sequence 5184, Ap
2	391	31.6	462	4 US-09-248-796A-5186	Sequence 5186, Ap
3	96.6	7.8	2320	4 US-09-016-434-1054	Sequence 1054, Ap
4	90.8	7.3	3798	4 US-09-949-016-4204	Sequence 4204, Ap
5	90.8	7.3	156942	4 US-09-949-016-12227	Sequence 12227, A
6	90.8	7.3	156950	4 US-09-949-016-15946	Sequence 15946, A
7	89.2	7.2	3839	4 US-09-949-016-485	Sequence 485, App
8	86.8	7.0	601	4 US-09-949-016-27359	Sequence 27359, A
9	86.8	7.0	601	4 US-09-949-016-149638	Sequence 149638, A
10	86.2	7.0	601	4 US-09-949-016-44799	Sequence 44799, A
11	86.2	7.0	45138	4 US-09-949-016-13027	Sequence 13027, A
12	85.6	6.9	601	4 US-09-949-016-27360	Sequence 27360, A
13	85.6	6.9	601	4 US-09-949-016-149637	Sequence 149637, A
14	83.2	6.7	601	4 US-09-949-016-27361	Sequence 27361, A
15	83.2	6.7	601	4 US-09-949-016-149636	Sequence 149636, A
16	81.6	6.6	472	4 US-09-513-899C-12522	Sequence 12522, A
17	80.2	6.5	2804	4 US-09-949-016-14278	Sequence 14278, Ap
18	80.2	6.5	22294	4 US-09-949-016-14020	Sequence 14020, A
19	79.4	6.4	1030	4 US-09-976-594-370	Sequence 370, App
20	79.2	6.4	2982	4 US-09-949-016-4184	Sequence 4184, App
21	79.2	6.4	41435	4 US-09-949-016-15926	Sequence 15926, A
22	78.2	6.3	3400	4 US-09-799-451-136	Sequence 136, App
23	77.2	6.2	601	4 US-09-949-016-149473	Sequence 149473, A
24	76	6.1	601	4 US-09-949-016-149474	Sequence 149474, A
25	76	6.1	2501	4 US-09-949-016-5363	Sequence 5363, App
26	76	6.1	2504	4 US-09-949-016-487	Sequence 487, App
27	76	6.1	29671	4 US-09-949-016-12229	Sequence 12229, A

28	76	6.1	29671	4 US-09-949-016-17105	Sequence 17105, A
29	74.4	6.0	2143	4 US-09-881-578A-3	Sequence 3, Appl
30	73	5.9	2925	4 US-09-620-312D-163	Sequence 163, App
31	71.2	5.7	979	4 US-09-976-594-939	Sequence 939, App
32	69.6	5.6	2402	4 US-09-949-016-1710	Sequence 1710, App
33	69.6	5.6	2407	4 US-09-949-016-489	Sequence 489, App
34	69.6	5.6	30867	4 US-09-949-016-12331	Sequence 12331, A
35	69.6	5.6	30868	4 US-09-949-016-13452	Sequence 13452, A
36	68	5.5	2891	4 US-09-949-016-1115	Sequence 1115, App
37	67.6	5.5	601	4 US-09-949-016-39838	Sequence 39838, A
38	67.6	5.5	601	4 US-09-949-016-160861	Sequence 160861, A
39	66.8	5.4	18773	4 US-09-949-016-14164	Sequence 14164, A
40	66.4	5.4	2891	4 US-09-949-016-4503	Sequence 4503, App
41	66.4	5.4	21360	4 US-09-949-016-12857	Sequence 12857, A
42	66.4	5.4	21361	4 US-09-949-016-16245	Sequence 16245, A
43	66.2	5.3	2752	4 US-09-949-016-3563	Sequence 3563, App
44	66.2	5.3	18824	4 US-09-949-016-15305	Sequence 15305, A
45	65.4	5.3	2274	4 US-09-620-312D-154	Sequence 154, App

ALIGNMENTS

RESULT 1
US-09-248-796A-5184
; Sequence 5184, Application US/09248796A
; Patent No. 6747137
; GENERAL INFORMATION:
; APPLICANT: Keith Weinstein et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBICAN
; FILE REFERENCE: 107196, 132
; CURRENT APPLICATION NUMBER: US/09/248, 796A
; CURRENT FILING DATE: 1999-02-12
; PRIOR APPLICATION NUMBER: US 60/074, 725
; PRIOR FILING DATE: 1998-02-13
; PRIOR APPLICATION NUMBER: US 60/096, 409
; PRIOR FILING DATE: 1998-08-13
; NUMBER OF SEQ ID NOS: 28208
; SEQ ID NO 5184
; LENGTH: 627
; TYPE: DNA
; ORGANISM: Candida albicans
US-09-248-796A-5184

Query Match	Score	DB 4	Length
Best Local Similarity	97.4%	Pred. No. 2.5e-124	
Matches 608; Conservative	0; Mismatches 7; Indels 9; Gaps 1;		
QY 616 TGTGTTGGGAAAAAGGTTTATCTTCACTATGTTAAGTCATGATGATTTCTACCATGATC	47.2%	627;	
DB 13 TGTGTTGGGAAAAAGGTTTATCTTCACTATGTTAAGTCATGATGATTTCTACCATGATC			72
QY 676 AAAATATGAGCTTGATATTGTGATGCGGAAATTTGCAAAAGAAATGAATTAGTT			725
DB 73 AAAATATGAGCTTGATATTGTGATGCGGAAATTTGCAAAAGAAATGAATTAGTT			132
QY 736 GAACATTATTAATCTTCATGATGATGTAATATCCCTGATGATTTATTAAGAACTGAA			795
DB 133 GAACATTATTAATCTTCATGATGATGTAATATCCCTGATGATTTATTAAGAACTGAA			192
QY 796 GTGAAAAAATTAGAAACCTATTAGTCAAGATGCAAAATTAATTTGCTGTAATTA			855
DB 193 GTGAAAAAATTAGAAACCTATTAGTCAAGATGCAAAATTAATTTGCTGTAATTA			252
QY 856 GAAACGAGAAATTAAAGTGAAGATGAAGAAATGAGATAGTCTAGATGAA			915
DB 253 GAAACGAGAAATTAAAGTGAAGATGAAGAAATGAGATAGTCTAGATGAA			303
QY 916 AAAAGAGTGATGTTAGATCAATGCTCAATGCTCAAGATCAATTAATCATTTACT			975
DB 304 AAAAGAGTGATGTTAGATCAATGCTCAATGCTCAAGATCAATTAATCATTTACT			363

QY 976 GCTTCTTTGGAGGTTCCAAAGAGTGTCTTAACTTATCTGTAATAGTGGAGAGATC 1035
| | | | |
DB 364 GCTTCTTTGGAGGTTCCAAAGAGTGTCTTAACTTATCTGTAATAGTGGAGAGATC 423
| | | | |
QY 1036 AATGTCTTGAAGATTAATGTGTAGTAAGTGTCTTGAAGATTAATAGTGGAGAGATC 1095
| | | | |
DB 424 AATGTCTTGAAGATTAATGTGTAGTAAGTGTCTTGAAGATTAATAGTGGAGAGATC 483
| | | | |
QY 1096 TTGAAATGGCATGATGATATTAATTAACAAGATTAATGATGATCTTAATAGTGAAGAAA 1155
| | | | |
DB 484 TTGAAATGGCATGATGATATTAATTAACAAGATTAATGATGATCTTAATAGTGAAGAAA 543
| | | | |
QY 1156 GAAGAAATCTCCAGAGGTGACCATTTGGTTAAAAAGCCAGATGATTTATGCCAAAT 1215
| | | | |
DB 544 GAAGAAATCTCCAGAGGTGACCATTTGGTTAAAAAGCCAGATGATTTATGCCAAAT 603
| | | | |
QY 1216 GAAACATCAGTGTCTTCTGATTA 1239
| | | | |
DB 604 GAAACATCAGTGTCTTCTGATTA 627
| | | | |

RESULT 2

US-09-248-796A-5186
; Sequence 5186, Application US/09248796A
; Patent No. 6747137
; GENERAL INFORMATION:
; APPLICANT: Keith Weinstock et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBICAN
; TITLE OF INVENTION: FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.132
; CURRENT APPLICATION NUMBER: US/09/248, 796A
; PRIOR FILING DATE: 1999-02-12
; PRIOR APPLICATION NUMBER: US 60/074, 725
; PRIOR FILING DATE: 1998-02-13
; PRIOR APPLICATION NUMBER: US 60/096, 409
; PRIOR FILING DATE: 1998-08-13
; NUMBER OF SEQ ID NOS: 28208
; SEQ ID NO 5186
; LENGTH: 462
; TYPE: DNA
; ORGANISM: Candida albicans
US-09-248-796A-5186

Query Match 31.6%; Score 391; DB 4; Length 462;
Best Local Similarity 98.5%; Pred. No. 4.4e-80;
Matches 405; Conservative 0; Mismatches 5; Indels 1; Gaps 1;

QY 1 ATGAGTGAAGTGAAGCAAAACCAATGATATCATCTTTAATCTTCTTCTTCATCA 60
| | | | |
DB 52 ATGAGTGAAGTGAAGCAAAACCAATGATATCATCTTTAATCTTCTTCTTCATCA 111
| | | | |
QY 61 CGTCCCAAAAAGATTTTGGCATATGAAAGGTGTGAATGAAGCCATATGACCATCA 120
| | | | |
DB 112 CGTCCCAAAAAGATTTTGGCATATGAAAGGTGTGAATGAAGCCATATGACCATCA 171
| | | | |
QY 121 TTATTAGCAACATTTAAGAAACCAAGTAAATGATGACCGTAAATAGTACAGTGAC 180
| | | | |
DB 172 TTATTAGCAACATTTAAGAAACCAAGTAAATGATGACCGTAAATAGTACAGTGAC 231
| | | | |
QY 181 GATGTGATTAAGCATTTTTCAGAAAATCACTTTGAAAACATATGTATCATCTTC 240
| | | | |
DB 232 GATGTGATTAAGCATTTTTCAGAAAATCACTTTGAAAACATATGTATCATCTTC 291
| | | | |
QY 241 GAAAAAACCATTGATGTTCAAGTGTGTTAAAGGGGTAAATCTGCAACAACCTTG 300
| | | | |
DB 292 GAAAAAACCATTGATGTTCAAGTGTGTTAAAGGGGTAAATCTGCAACAACCTTG 351
| | | | |
QY 301 AAAAGCATGAATCAACCATACAAAGTCAATTAATGTAAATTTGAAATTTGTCAAGA 360
| | | | |
DB 352 AAAAGCATGAATCAACCATACAAAGTCAATTAATGTAAATTTGAAATTTGTCAAGA 411
| | | | |
QY 361 GCATTTTAT-AAACATCAATCTTTAAGACATATATATCTGTCATGA 410
| | | | |

DB 412 ACATTTTATAAAACATCAATCTTTAAGACATCATATATATCTGTCATGA 462
| | | | |

RESULT 3

US-09-016-434-1054
; Sequence 1054, Application US/09016434
; Patent No. 6500938
; GENERAL INFORMATION:
; APPLICANT: Janice Au-Young
; APPLICANT: Jeffrey J. Seilhamer
; TITLE OF INVENTION: COMPOSITION FOR THE DETECTION OF SIGNALING
; TITLE OF INVENTION: PATHWAY GENE EXPRESSION
; NUMBER OF SEQUENCES: 1490
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: INCYTE PHARMACEUTICALS, INC.
; STREET: 3174 PORTER DRIVE
; CITY: PALO ALTO
; STATE: CALIFORNIA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Word Perfect 6.1 for Windows/MS-DOS 6.2
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/016, 434
; FILING DATE: HERewith
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Zeller, Karen J.
; REGISTRATION NUMBER: 37,071
; REFERENCE/DOCKET NUMBER: PA-0002 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (650) 855-0555
; TELEFAX: (650) 845-4166
; INFORMATION FOR SEQ ID NO: 1054:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2320 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; LIBRARY: GENBANK
; CLONE: 91017721
US-09-016-434-1054

Query Match 7.8%; Score 96.6; DB 4; Length 2320;
Best Local Similarity 51.3%; Pred. No. 1.4e-12;
Matches 306; Conservative 0; Mismatches 279; Indels 12; Gaps 3;

QY 69 AAGATATATTTGGCATATGAAAGGTGTGATTAAGCCATATATGACCATCATTTATGA 128
| | | | |
DB 964 AAGAAACCCATCAAAATGTAAAGATGTGTAAAGCTTTAAACGATCTTCAACCTTAC 1023
| | | | |
QY 129 GCAACATTTAAGAACCCACAGTATGATGACCGTATTAATGTATACAGTGGAGATGTGA 188
| | | | |
DB 1024 TACCATGAAAAAATTTACTGAGAGAAACCTTACAAATGT-----GAAAGATGTGG 1077
| | | | |
QY 189 TAAACATTTTTCAGAAAATCACTTTGAAAACATATGTATATCATCTTCGGAAGAAA 248
| | | | |
DB 1078 CAAGCCCTTTAAGCAGTCTCTCAACCTTACATACATTAAGATTAATCTACTGAGAGAA 1137
| | | | |
QY 249 ACCATTCATGTTCAAGTGTGTGAAGGGGTAAATCTGCAACAACCTTGAAGAAACA 308
| | | | |
DB 1138 ACCCTACAAATGTAAATATGTGAAGCCCTTTAACCAGTCTGACACCTTACACACA 1197
| | | | |
QY 309 TGAATCAACCATCAAGATGATTAATGTATCAATTTGAAAATGTGCAAGAGATTTTA 368
| | | | |


```

Db      1198 TGAAGTAATTCATCTGAGAGAAACCTTACAATGTGAAAAATGTGAAAAACCTTTAA 1257
Qy      369 TAAACA---TCATCTTTAAGACATCATATATATGTTTCATGAAAAAATTAACGCG 425
Db      1258 TCAATTTCTACACCTTACTACATAGAAATTTCAATCTGAGAGAAACCTTACAAAG 1317
Qy      426 TAAACATGTATTAAGATTTTCACTGACCTTCAAAATTTAGACAACATTAATTTAAAGA 485
Db      1318 TAAAGATGTGTAAGACCTTTTAAACATCTTCAACCTTACAAATTAAGTATTTCA 1377
Qy      486 TCATGTGTATCTCTGCTTATCATATGTATCATCTGTTTAAAAATTTCCAAAC 545
Db      1378 TACTGGAGAGAAAGCTTACAAATCTAAGATGTGAAAAAGCTTTAACCAATCTCTCAA 1437
Qy      546 TTGCTGATTTTCAATTTTCAATTAACCACTGATCCAAAATTAATGTCTTAATG 605
Db      1438 ACTTACTGAACATTAAGAAATTTATCTGAGAGAAACCTTA---TGAATGTGAAAAATG 1494
Qy      606 TGGTAAGGTGTGTGGGAAAAAGTTTATCTTCAATATGTGTAAGTATGATGA 662
Db      1495 TGGCAAGCTTTTAAACCAAGCTTCAATCTTACTAAGACATTAAGAAAGTATACAGA 1551

RESULT 4
US-09-949-016-4204
; Sequence 4204, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 4204
; LENGTH: 3798
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-4204

Query Match
Best Local Similarity 7.3%; Score 90.8; DB 4; Length 3798;
Matches 235; Conservative 0; Mismatches 192; Indels 9; Gaps 2;

Qy      69 AAAGTATATTTGCATATGAAAGGTGTGATTAAGCCTTAATTCAGACATCATTTATTA 128
Db      1848 AGGAAACCTTACAAATGTGAAGATGTGCAAAAGCTTTAGCATTTCTTCAACCTTGC 1907
Qy      129 GCAACATTTAAGAACCCACAGTATGATCGACCGTAAATTTAGACAGTGGACGATTGTA 188
Db      1908 TAAACATTAAGAAATTCATCTGAGAGAAACCTTCAAAATGT-----GAAGAAATGTG 1961
Qy      189 TAAAGATTTTGAAGAAATTCATTTGGAACAACATATGTATCATCTTCCGAAAAAAA 248
Db      1962 CAAAGCTTTAGCGCTTCTTCAACCTTGTAAACATTAAGAAATTCATCTGAGAGAA 2021
Qy      249 ACCATTCATTTGTCTAGTGTGTGTAAGAGGTTAAATTCGACAAACATTTGAAAAAGA 308
Db      2022 ACCCTCAATGTAAAGAAATGTGCAAAAGCTTTTGAATTCCTCAACCTTGTATATCA 2081
Qy      309 TGAATTCACCCATCAAAAGTCAATTTAATGTATCAATTTGAAATTTGCAAGACATTTTA 368
Db      2082 TAAAGATTAATCTCATAGAGAGAAACCTTCAAAATGTAAAGAAATGTGACAAAACTTTTA 2141
Qy      369 TAAACATCATC---TTTAAAGATCATATATATCTGTTCAAGAAAAAATTAACGCG 425

```

```

Db      2142 GGACCTCTCAACCTTACTTAACATTAATTAATACATGCTGAGAGAAACCTTACAAAG 2201
Qy      426 TAAACATGTATTAAGATTTTCACTGACCTTCAAAATTTAGACAACATTAATTAAGA 485
Db      2202 TGAAGATGTGCAAAAGCTTTTAAATGTATGATCTTCAATCTTACTATACATTAAGTTTATCA 2261
Qy      486 TCATGTGTATCTCTGCTTATCATATGTATCATCTGTTTAAAAATTTCCAAAC 501
Db      2262 TACTGGAGAGAAACCT 2277

RESULT 5
US-09-949-016-12227
; Sequence 12227, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 12227
; LENGTH: 156942
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)...(156942)
; OTHER INFORMATION: n = A,T,C or G
US-09-949-016-12227

Query Match
Best Local Similarity 7.3%; Score 90.8; DB 4; Length 156942;
Matches 235; Conservative 0; Mismatches 192; Indels 9; Gaps 2;

Qy      69 AAAGTATATTTGCATATGAAAGGTGTGATTAAGCCTTAATTCAGACATCATTTATTA 128
Db      152999 AGGAAACCTTACAAATGTGAAGATGTGCAAAAGCTTTAGCATTTCTTCAACCTTGC 153058
Qy      129 GCAACATTTAAGAACCCACAGTATGATGACCGTAAATTTAGACAGTGGACGATTGTA 188
Db      153059 TAAACATTAAGAAATTCATCTGAGAGAAACCTTCAAAATGT-----GAAGAAATGTG 153112
Qy      189 TAAAGATTTTGAAGAAATTCATTTGGAACAACATATGTATCATCTTCCGAAAAAAA 248
Db      153113 CAAAGCTTTAGCGCTTCTTCAACCTTGTAAACATTAAGAAATTCATCTGAGAGAA 153172
Qy      249 ACCATTCATTTGTCTAGTGTGTGTAAGAGGTTAAATTCGACAAACATTTGAAAAAGA 308
Db      153173 ACCCTCAATGTAAAGAAATGTGCAAAAGCTTTTGAATTCCTCAACCTTGTATATCA 153232
Qy      309 TGAATTCACCCATCAAAAGTCAATTTAATGTATCAATTTGAAATTTGCAAGACATTTTA 368
Db      153233 TAAAGATTAATCTCATAGAGAGAAACCTTCAAAATGTAAAGAAATGTGACAAAACTTTTA 153292
Qy      369 TAAACATCATC---TTTAAAGATCATATATATCTGTTCAAGAAAAAATTAACGCG 425
Db      153293 GGAATCTTCAACCTTACTTAACATTAATTAATATCATGCTGAGAGAAACCTTACAAATG 153352
Qy      426 TAAACATGTATTAAGATTTTCACTGACCTTCAAAATTTAGACAACATTAATTAAGA 485
Db      153353 TGAAGATGTGCAAAAGCTTTTAAATGTATGATCTTCAATCTTACTATACATTAAGTTTATCA 153412

```

QY 486 TCATGGTGAATCTCCT 501
DB 153413 TACTGGAGAGAAACCT 153428

RESULT 6

US-09-949-016-15946
Sequence 15946, Application US/09949016
Patent No. 6812339
GENERAL INFORMATION:
APPLICANT: VENTER, J. Craig et al.
TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
FILE REFERENCE: CL001307
CURRENT FILING DATE: 2000-04-14
PRIOR APPLICATION NUMBER: 60/241,755
PRIOR FILING DATE: 2000-10-20
PRIOR APPLICATION NUMBER: 60/237,768
PRIOR FILING DATE: 2000-10-03
PRIOR APPLICATION NUMBER: 60/231,498
PRIOR FILING DATE: 2000-09-08
NUMBER OF SEQ ID NOS: 207012
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 15946
LENGTH: 156950
TYPE: DNA
ORGANISM: Human
FEATURE:
NAME/KEY: misc_feature
LOCATION: (1)...(156950)
OTHER INFORMATION: n = A,T,C or G
US-09-949-016-15946

Query Match 7.3%; Score 90.8; DB 4; Length 156950;
Best Local Similarity 53.9%; Pred. No. 1.1e-10;
Matches 235; Conservative 0; Mismatches 192; Indels 9; Gaps 2;

QY 69 AAGATATATTTGACATATGAGGGGTGATTAAGGCTTAATTCAGCATCTATTATTAGA 128
DB 152999 AGAGAAACCTTACAAATGTGAAGATGTGGCAAGCTTTTAGCATCTTCAACCTTGC 153058
QY 129 GCAACATTTAAGAACCCACAGTAATGATGACCGTAAATGTACAGTGAAGATTGTGA 188
DB 153059 TAAACATAGAGAAATCATACGAGAGAAACCTTCAATGT-----GAAGATGTGG 153112
QY 189 TAAAGATTTTCAAGAAATCACTTTGAAACACATATTGATCATCTCCGAAAAA 248
DB 153113 CAAAGCTTTAGCGCTTCTTCAACCTTGTCTAAACATAAGAAATTCATACCTGAGAGAA 153172
QY 249 ACCATTCATTTGTCAGTGTGTGTAAGGGGTTAATTCGACACACTTGAAGAACA 308
DB 153173 ACCCTCAATGTGAAGATGTGGCAAGCTTTTAGCATCTTCAACCTTGTCTATCA 153232
QY 309 TGAATACCCATACAAAGTCAATTAATGTACATTTGAAAAATTGTCAAGAACATTTTA 368
DB 153233 TAAAGATTACTCACTAGAGAGAAACCTTCAAAATGTAAGATGTGACAAAACTTTTA 153292
QY 369 TAAACATCAATC---TTTAAACATCATATATTATCTGTCATGAAAAAATTAACGTG 425
DB 153293 GCGACTCTCAACCTTACCTTAACATTAATTAATTAATTAATTAATTAATTAATTA 153352
QY 426 TAAACATGTAATTAAGTTTCACTGACCTTCAAAATTAAGACACATTAATTAATAACA 485
DB 153353 TGAAGATGTGGCAAGCTTTTATGATCTTCAAAATCTTACTATACATTAATTAATTA 153412
QY 486 TCATGGTGAATCTCCT 501
DB 153413 TACTGGAGAGAAACCT 153428

RESULT 7
US-09-949-016-485

/ Sequence 485, Application US/09949016
/ Patent No. 6812339
/ GENERAL INFORMATION:
/ APPLICANT: VENTER, J. Craig et al.
/ TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
/ WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
/ FILE REFERENCE: CL001307
/ CURRENT FILING DATE: 2000-04-14
/ PRIOR APPLICATION NUMBER: 60/241,755
/ PRIOR FILING DATE: 2000-10-20
/ PRIOR APPLICATION NUMBER: 60/237,768
/ PRIOR FILING DATE: 2000-10-03
/ PRIOR APPLICATION NUMBER: 60/231,498
/ PRIOR FILING DATE: 2000-09-08
/ NUMBER OF SEQ ID NOS: 207012
/ SOFTWARE: FastSeq for Windows Version 4.0
/ SEQ ID NO 485
/ LENGTH: 3839
/ TYPE: DNA
/ ORGANISM: Human
US-09-949-016-485
Query Match 7.2%; Score 89.2; DB 4; Length 3839;
Best Local Similarity 53.7%; Pred. No. 8e-11;
Matches 234; Conservative 0; Mismatches 193; Indels 9; Gaps 2;

QY 69 AAGATATATTTGACATATGAGGGGTGATTAAGGCTTAATTCAGCATCTATTATTAGA 128
DB 964 AGAGAAACCTTACAAATGTGAAGATGTGGCAAGCTTTTAGCATCTTCAACCTTGC 1023
QY 129 GCAACATTTAAGAACCCACAGTAATGATGACCGTAAATGTACAGTGAAGATTGTGA 188
DB 1024 TAAACATAGAGAAATCATACCTGAGAGAAACCTTCAATGT-----GAAGATGTGG 1077
QY 189 TAAACATTTTCAAGAAATCACTTTGAAACACATATTGATCATCTCCGAAAAA 248
DB 1078 CAAAGCTTTAGCGCTTCTTCAACCTTGTCTAAACATTAAGAAATTCATACCTGAGAGAA 1137
QY 249 ACCATTCATTTGTCAGTGTGTGTAAGGGGTTAATTCGACACACTTGAAGAACA 308
DB 1138 ACCCTCAATGTGAAGATGTGGCAAGCTTTTAGCATCTTCAACCTTGTCTATCA 1197
QY 309 TGAATACCCATACAAAGTCAATTAATGTACATTTGAAAAATTGTCAAGAACATTTTA 368
DB 1198 TAAAGATTACTCACTAGAGAGAAACCTTCAAAATGTAAGATGTGACAAAACTTTTA 1257
QY 369 TAAACATCAATC---TTTAAACATCATATATTATCTGTCATGAAAAAATTAACGTG 425
DB 1258 GCGACTCTCAACCTTACCTTAACATTAATTAATTAATTAATTAATTAATTAATTA 1317
QY 426 TAAACATGTAATTAAGTTTCACTGACCTTCAAAATTAAGACACATTAATTAATAACA 485
DB 1318 TGAAGATGTGGCAAGCTTTTATGATCTTCAAAATCTTACTATACATTAATTAATTA 1377
QY 486 TCATGGTGAATCTCCT 501
DB 1378 TACTGGAGAGAAACCT 1393

RESULT 8
US-09-949-016-27359/c
Sequence 27359, Application US/09949016
Patent No. 6812339
GENERAL INFORMATION:
APPLICANT: VENTER, J. Craig et al.
TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
FILE REFERENCE: CL001307
CURRENT FILING DATE: 2000-04-14
PRIOR APPLICATION NUMBER: 60/241,755
PRIOR FILING DATE: 2000-10-20

```

; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 27359
; LENGTH: 601
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-27359

```

```

Query Match
Best Local Similarity 51.4%; Pred. No. 1.6e-10;
Matches 252; Conservative 1; Mismatches 228; Indels 9; Gaps 2;

```

```

QY 7 GAAAGTGACGAAACCAATGATATCATCTTAAATCTTCTTCTTCATCAGTCCC 66
DB 538 GAATGTGGCAAAAGCTTTTAAAGCACTCCTCAGCCCTTGCTAAACATTAATATCATGTCT 479
QY 67 AAAAGTATATTTGCAATATGAAAGGTGATTAAGCCCTAATGACATCATATTA 126
DB 478 GGAAGAACTTACAAATGTGAGAAATGTGCAAAAGCTTTAATCATCTTCAAAATCTT 419
QY 127 GAGCAACATTTAAGAACCCACAGTAATGATCGACCGTAAATGTAACAGTGAAGATGT 186
DB 418 AGACACATTAAGTAATTAATCTACTTAAGAAACCTTCCAGAGT-----GAAAGATGT 365
QY 187 GATTAAGCATTTTTCAGAAAAATCATATTGAAAAACATATTGTATCATCTCCGAAAA 246
DB 364 GACAAAGCATTTTATCTGTCTCTCAACCTTACTGAACATTAAGAAATTCATACAGAGAG 305
QY 247 AAACCATTCATTTGTCAGTGTGTGTAAGGGGTTAATCTGACAAACCTTGAAGA 306
DB 304 AAACCTCAAAATGTGAAAGATGTGCAAAAGCATTTTACGAGCCTTCACTTACTACA 245
QY 307 CATGAATACCCATACAAAGTCAATTTAATGTACATTTGAAATTTGCAAGAGATTT 366
DB 244 CATTAAGAGATGACACTGAGAGAAACCTTACAAATGTGAAGAAATGTGCAAAAGCTTTT 185
QY 367 TATTAACATCATATC--TTTAAGACATCATATTAATCTGTTGATGAAAAACATTAAG 423
DB 184 AGCCAAATCCTCAACCTTACTACTACATTAAGATTAATCATCTGAGAGAAACCTTCAAA 125
QY 424 TGTAAACATGTATTAAGTTTCACTGACCTTCAAAATTAAGCAACATTAATTA 483
DB 124 TGTGAAGAAATGTGCAAAAGCTTTTGAAGAAATCTTCACTGAAACATTAAGATTAAT 65
QY 484 CATCATGTGTG 493
DB 64 CATACTGGAG 55

```

```

RESULT 9
US-09-949-016-149638/c
; Sequence 149638, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; PRIOR FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 149638

```

```

; LENGTH: 601
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-149638

```

```

Query Match
Best Local Similarity 51.4%; Pred. No. 1.6e-10;
Matches 252; Conservative 1; Mismatches 228; Indels 9; Gaps 2;

```

```

QY 7 GAAAGTGACGAAACCAATGATATCATCTTAAATCTTCTTCTTCATCAGTCCC 66
DB 538 GAATGTGGCAAAAGCTTTTAAAGCACTCCTCAGCCCTTGCTAAACATTAATATCATGTCT 479
QY 67 AAAAGTATATTTGCAATATGAAAGGTGATTAAGCCCTAATGACATCATATTA 126
DB 478 GGAAGAACTTACAAATGTGAGAAATGTGCAAAAGCTTTAATCATCTTCAAAATCTT 419
QY 127 GAGCAACATTTAAGAACCCACAGTAATGATCGACCGTAAATGTAACAGTGAAGATGT 186
DB 418 AGACACATTAAGTAATTAATCTACTTAAGAAACCTTCCAGAGT-----GAAAGATGT 365
QY 187 GATTAAGCATTTTTCAGAAAAATCATATTGAAAAACATATTGTATCATCTCCGAAAA 246
DB 364 GACAAAGCATTTTATCTGTCTCTCAACCTTACTGAACATTAAGAAATTCATACAGAGAG 305
QY 247 AAACCATTCATTTGTCAGTGTGTGTAAGGGGTTAATCTGACAAACCTTGAAGA 306
DB 304 AAACCTCAAAATGTGAAAGATGTGCAAAAGCATTTTACGAGCCTTCACTTACTACA 245
QY 307 CATGAATACCCATACAAAGTCAATTTAATGTACATTTGAAATTTGCAAGAGATTT 366
DB 244 CATTAAGAGATGACACTGAGAGAAACCTTCAAAATGTGAAGAAATGTGCAAAAGCTTTT 185
QY 367 TATTAACATCATATC--TTTAAGACATCATATTAATCTGTTGATGAAAAACATTAAG 423
DB 184 AGCCAAATCCTCAACCTTACTACTACATTAAGATTAATCATCTGAGAGAAACCTTCAAA 125
QY 424 TGTAAACATGTATTAAGTTTCACTGACCTTCAAAATTAAGCAACATTAATTA 483
DB 124 TGTGAAGAAATGTGCAAAAGCTTTTGAAGAAATCTTCACTGAAACATTAAGATTAAT 65
QY 484 CATCATGTGTG 493
DB 64 CATACTGGAG 55

```

```

RESULT 10
US-09-949-016-44799
; Sequence 44799, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; PRIOR FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 44799
; LENGTH: 601
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-44799

```

```

Query Match
Best Local Similarity 54.4%; Pred. No. 2.2e-10;

```

Matches 241; Conservative 0; Mismatches 193; Indels 9; Gaps 3;

QY 61 CGTCCCAAAAGATATTTGCAATATGAGGGTGTATMAAGCCTATATGACCATCA 120
DB 56 CATCTAGAGAGAAACCTTACAAATGTGAAGATGTGGCAATTTATATGTCTCTCA 115
QY 121 TTATTAGACACATTTAAGAACCCACAGTAATGATCGACCGTATTAATGTACAGTGAC 180
DB 116 ACCCTAACTAGACATTAAGAGATGACACTGAGAGAAACCCCTACAAATGT-----GAA 169
QY 181 GATTGTGATTAAGACATTTTTCAGAAATGACATTTGGAACATATGTATCATCATTC 240
DB 170 GATATGGCAATCTTTTAGCCATCTCTACACCTTACTACACATTAAGATATTCATCT 229
QY 241 GAAAAAAACCATTCATTGTCAGTGTGTGTGTAAGGGGTAAATTCGACAAACCTTG 300
DB 230 GGAGAGAAACCTTACAAATGTGAAGATGTGGCAAGCCTTTAACTGTCTCAACTCTT 289
QY 301 AAAAGACATGAATGACCCATACAAAGTCATTTAAATGTACATTTGAAATTTGTCAAGA 360
DB 290 ACTAAACATTAATATTTCTACTGAGAGAAACCCCTACAAATGTGAAAAATGTGGCAAA 349
QY 361 GCATTTTA--TAAACATCAATCTTTA--AGACATCATATATATCTGTGATGAAAAACA 417
DB 350 GCCTTTAGACAGCTTCAATCTTACTACTACATTAAGAGATTCATCTGAGAGAAACCC 409
QY 418 TTAACGTGTAAACAATGATTAATAAGTTTCACTGACCTTCAAATTTAGCACAACATAA 477
DB 410 TATAAATGTGAAGATGTGGCAAACTTTTAAACCGGTCTTCAACTTTTCTAAACATAAG 469
QY 478 TTTAAACATCATGTGTGATCTCC 500
DB 470 GTAATTCATCTAGAGTAAACCC 492

RESULT 11
US-09-949-016-13027
Sequence 13027, Application US/09949016
Patent No. 6812339
GENERAL INFORMATION:
APPLICANT: VENTER, J. Craig et al.
TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
FILE REFERENCE: CL001307
CURRENT APPLICATION NUMBER: US/09/949,016
PRIOR FILING DATE: 2000-04-14
PRIOR APPLICATION NUMBER: 60/241,755
PRIOR FILING DATE: 2000-10-20
PRIOR APPLICATION NUMBER: 60/237,768
PRIOR FILING DATE: 2000-10-03
PRIOR APPLICATION NUMBER: 60/231,498
PRIOR FILING DATE: 2000-09-08
NUMBER OF SEQ ID NOS: 207012
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 13027
LENGTH: 45138
TYPE: DNA
ORGANISM: Human
FEATURE:
NAME/KEY: misc_feature
LOCATION: (1)...(45138)
OTHER INFORMATION: n = A,T,C or G
US-09-949-016-13027

Query Match 7.0%; Score 86.2; DB 4; Length 45138;
Best Local Similarity 54.4%; Pred. No. 8.4e-10;
Matches 241; Conservative 0; Mismatches 193; Indels 9; Gaps 3;

QY 61 CGTCCCAAAAGATATTTGCAATATGAGGGTGTATMAAGCCTATATGACCATCA 120
DB 42121 CATCTAGAGAGAAACCTTACAAATGTGAAGATGTGGCAAGCATTATATGTCTCTCA 42180
QY 121 TTATTAGACACATTTAAGAACCCACAGTAATGATCGACCGTATTAATGTACAGTGAC 180

DB 42181 ACCCTAACTAGACATTAAGAGATGCACTGAGAGAAACCCCTACAAATGT-----GAA 42234
QY 181 GATTGTGATTAAGACATTTTTCAGAAATGACATTTGGAACATATGTATGATTCATTC 240
DB 42235 GAAATGGCAAACTTTTAGCCATCTCTACACCTTACTACACATTAAGATATTCATCT 42294
QY 241 GAAAAAAACCATTCATTGTCAGTGTGTGTGTAAGGGGTAAATTCGACAAACCTTG 300
DB 42295 GGAGAGAAACCTTACAAATGTGAAGATGTGGCAAGCCTTTAACTGTCTCTCAACTCTT 42354
QY 301 AAAAGACATGAATGACCCATACAAAGTCATTTAAATGTACATTTGAAATTTGTCAAGA 360
DB 42355 ACTTAACATTAAGATATTTCTACTGAGAGAAACCCCTCAAAATGTGAAAAATGTGGCAAA 42414
QY 361 GCATTTTA--TAAACATCAATCTTTA--AGACATCATATATATCTGTTCATGAAAAACA 417
DB 42415 GCCTTTAGACAGCTTCAATCTTACTACTACATTAAGAGATTCATCATCTGAGAGAAACCC 42474
QY 418 TTAACGTGTAAACAATGATTAATAAGTTTCACTGACCTTCAAATTTAGCACAACATAA 477
DB 42475 TATAAATGTGAAGATGTGGCAAACTTTTAAACCGGTCTTCAACTTTTACTTAACATTAAG 42534
QY 478 TTTAAACATCATGTGTGATCTCC 500
DB 42535 GTAATTCATCTAGAGTAAACCC 42557

RESULT 12
US-09-949-016-27360/C
Sequence 27360, Application US/09949016
Patent No. 6812339
GENERAL INFORMATION:
APPLICANT: VENTER, J. Craig et al.
TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
FILE REFERENCE: CL001307
CURRENT APPLICATION NUMBER: US/09/949,016
PRIOR FILING DATE: 2000-04-14
PRIOR APPLICATION NUMBER: 60/241,755
PRIOR FILING DATE: 2000-10-20
PRIOR APPLICATION NUMBER: 60/237,768
PRIOR FILING DATE: 2000-10-03
PRIOR APPLICATION NUMBER: 60/231,498
PRIOR FILING DATE: 2000-09-08
NUMBER OF SEQ ID NOS: 207012
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 27360
LENGTH: 601
TYPE: DNA
ORGANISM: Human
US-09-949-016-27360

Query Match 6.9%; Score 85.6; DB 4; Length 601;
Best Local Similarity 51.4%; Pred. No. 3e-10;
Matches 252; Conservative 0; Mismatches 229; Indels 9; Gaps 2;

QY 7 GAAAGTGAAGAAACCAATGATATCATCTTTATATATCTCTCTCTTCATGACGTC 66
DB 499 GAAATGGCAAAAGCTTTTAAGACATCTCTAGCCCTGTCTTAACAATATATATCATGT 440
QY 67 AAAAGATATTTGCAATATGAGGGTGTGATTAAGCCTATATGACCATCATTAATTA 126
DB 439 GGAGAGAAACCTTACAAATGTGAAGATGTGGCAAGCCTTTTAATCAATCTTCAATCTT 380
QY 127 GAGCAACATTTTAAGAACCCACAGTAATGATGACCGGTATTAATGTACAGTGACATGT 186
DB 379 ACGACACATTAAGATATTTCTACTAAGAGAAACCTTCCAAAGAT-----GAAAGATGT 326
QY 187 GATAAAGATTTTTCAGAAATGACATTTGGAACACATATTTGTATTCATTCGAAAAA 246
DB 325 GACAAAGCATTTATCTGTCTCTACACCTTACTGAAACATTAAGAGATTCATTCAGAGAG 266

QY 247 AAACATTCATTGTCAGTGTGTGTAAGGGGTTAATTCGACAAACATTGAAAGA 306
DB 265 AAACCTCAAAATGTGAAGATGTGGCAAGCATTTTACGAGCTTACACTTACACA 206
QY 307 CATGAATTCACCATACAAAGTCATTTAAATGTACATTTGAAAATGTCAAGACATTT 366
DB 205 CATGAAGAGTGTGCAACTGTGAGAGAAACCTACAAATGTGAAATGTGCAAGCTTTT 146
QY 367 TATAAATCATCAATC---TTTAAGACATCATATATTAATCTGTGATGAAAACATTAACG 423
DB 145 AGCCAAATCCTCAACCTTACTACATTAAGATTAATTCATCTGAGAGAAACCTTACAAA 86
QY 424 TGTAAACAATGTATAAAGTTTTCACCTGACCTTCAAAATTAACACATTAATTAATA 483
DB 85 TGTGAAGAATGTGGCAAAAGCTTTTGAAGAAATCTTCAACTTACTGAACTTAAGATTAAT 26
QY 484 CATCATGTGTG 493
DB 25 CATACTGGAG 16

RESULT 13
US-09-949-016-149637/c
; Sequence 149637, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; PRIOR FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 149637
; LENGTH: 601
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-149637

Query Match
Best Local Similarity 51.4%; Score 85.6; DB 4; Length 601;
Matches 252; Conservative 0; Mismatches 229; Indels 9; Gaps 2;

QY 7 GAAAGTGACAAACCAATGATATCATCTTTAATATCTTCTTCTTCATCAGTCCC 66
DB 499 GAATGTGGCAAAAGCTTTTAAAGCACTCCTCAGCCCTGTAAACATTAATAATACATGCT 440
QY 67 AAAAGATATTTGCAATATGAAGGGTGTGTAAGGCTTAATGAGCATTATTAATTA 126
DB 439 GGAAGAAATCTTCAAAATGTGAGAAATGTGGCAAGCTTTTAAATCTTAAATCTT 380
QY 127 GAGCAATTTTAAAGAACCCAGATTAATGATCGACCTATTAATGTACAGTGAAGTGT 186
DB 379 AGACACATTAAGTAATTTACTTAAGAGAAACCTTCCAGAGT-----GAAAGATGT 326
QY 187 GATTAAGCATTTTTCAGAAAATCAATTGGAAAACATATTGTATCACTTCGAAAA 246
DB 325 GACAAAGCATTTTATGTGTCCTCCTCAGCCCTTACGAAATTAAGAAATTTCAACAGAG 266
QY 247 AAACATTCATTGTCAGTGTGTGTAAGGGGTTAATTCGACAAACATTGAAAGA 306
DB 265 AAACCTCAAAATGTGAAGATGTGGCAAGCATTTTACGAGCTTACACTTACACA 206
QY 307 CATGAATTCACCATACAAAGTCATTTAAATGTACATTTGAAAATGTCAAGACATTT 366
DB 205 CATGAAGAGTGTGCAACTGTGAGAGAAACCTACAAATGTGAAATGTGCAAGCTTTT 146

QY 367 TATAAATCATCAATC---TTTAAGACATCATATATTAATCTGTGATGAAAACATTAACG 423
DB 145 AGCCAAATCCTCAACCTTACTACATTAAGATTAATTCATCTGAGAGAAACCTTACAAA 86
QY 424 TGTAAACAATGTATAAAGTTTTCACCTGACCTTCAAAATTAACACATTAATTAATA 483
DB 85 TGTGAAGAATGTGGCAAAAGCTTTTGAAGAAATCTTCAACTTACTGAACTTAAGATTAAT 26
QY 484 CATCATGTGTG 493
DB 25 CATACTGGAG 16

RESULT 14
US-09-949-016-27361/c
; Sequence 27361, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; PRIOR FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 27361
; LENGTH: 601
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-27361

Query Match
Best Local Similarity 52.5%; Score 83.2; DB 4; Length 601;
Matches 231; Conservative 1; Mismatches 199; Indels 9; Gaps 2;

QY 57 ATCAGCTCCCAAAAGTATATTTGCAATATGAAGGGTGTGTAAGGCTTATATGACC 116
DB 554 AACTCATCTGGAAGAGAAACCTTACAAATGTAAAGATGTGACAAACCTTTAAGGACT 495
QY 117 ATCATATTAAGCAACATTTAAGAACCGACAGTATGATCGACGCTATTAATGTACAGT 176
DB 494 CTCACCTTACTTAAACATTAATTAATTAATCACTGTGAGAGAAACCTTACAAATGT----- 440
QY 177 GAGCATGTGTGTAAGCAATTTTTCAGAAAATCAATTGGAAAACATATTGTATCA 236
DB 439 -GAGAAATGTGGAAAGCTTTTAAATGATCTTCAATCTTACATTAAGTATTTTCA 381
QY 237 TTCGAAAAAAACCATTCATTTTCAAGTGTGTGTAAGGGGTTAATTTCTGACACA 296
DB 380 TACTGAGAGAAACCTTACAGGTGTGMAAATGTGCAAGCATTTTAACTGTCTCAAG 321
QY 297 CTGGAAGAAACATGAATACCCATCAAGATCATTAATTAATGTATGATTTGAAAATGTCA 356
DB 320 CCTTACTTAACATTAAGAAATTAATCTAGAGAGAAACCTTCAAAATGTGAAGATGTG 261
QY 357 AGAAGCATTTTATTAACATCAATCTTTA---AGACATCATATTAATTTCTGTCATGAAA 413
DB 260 CAAGCATTTTATATGTCTTCAACCTTACTGACATTAAGAGATCAACACTGAGAGAA 201
QY 414 AACATTAACGTGTAAACATGTATTAAGTTTTCATGACCTTCAAAATTAACACACA 473
DB 200 GCCCTTCAAAATGTGAAGATGTGGCAAGCTTTTACGCTTCTTCAACCTTACTTAAGCA 141
QY 474 TAAATTAACATCATGTGTG 493

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 9, 2005, 05:36:40 ; Search time 2160.11 Seconds
(without alignments)
3301.382 Million cell updates/sec

Title: US-09-831-804-2

Perfect score: 1239
Sequence: 1 atgagtgaaagtcagcaaac.....catcagctgattctcgataa 1239

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 4313806 seqs, 2877871033 residues

Total number of hits satisfying chosen parameters: 8627612

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

Published Applications NA:*

- 1: /cgn2_6/prodata/1/pubpna/US07_PUBCOMB.seq:*
- 2: /cgn2_6/prodata/1/pubpna/PCR_NEW_PUB.seq:*
- 3: /cgn2_6/prodata/1/pubpna/US06_NEW_PUB.seq:*
- 4: /cgn2_6/prodata/1/pubpna/US06_PUBCOMB.seq:*
- 5: /cgn2_6/prodata/1/pubpna/US07_NEW_PUB.seq:*
- 6: /cgn2_6/prodata/1/pubpna/PCRUS_PUBCOMB.seq:*
- 7: /cgn2_6/prodata/1/pubpna/US08_NEW_PUB.seq:*
- 8: /cgn2_6/prodata/1/pubpna/US08_PUBCOMB.seq:*
- 9: /cgn2_6/prodata/1/pubpna/US09_PUBCOMB.seq:*
- 10: /cgn2_6/prodata/1/pubpna/US09_PUBCOMB.seq:*
- 11: /cgn2_6/prodata/1/pubpna/US09C_PUBCOMB.seq:*
- 12: /cgn2_6/prodata/1/pubpna/US09C_NEW_PUB.seq:*
- 13: /cgn2_6/prodata/1/pubpna/US10_PUBCOMB.seq:*
- 14: /cgn2_6/prodata/1/pubpna/US10_PUBCOMB.seq:*
- 15: /cgn2_6/prodata/1/pubpna/US10C_PUBCOMB.seq:*
- 16: /cgn2_6/prodata/1/pubpna/US10_PUBCOMB.seq:*
- 17: /cgn2_6/prodata/1/pubpna/US10_PUBCOMB.seq:*
- 18: /cgn2_6/prodata/1/pubpna/US10_PUBCOMB.seq:*
- 19: /cgn2_6/prodata/1/pubpna/US10_NEW_PUB.seq:*
- 20: /cgn2_6/prodata/1/pubpna/US11_NEW_PUB.seq:*
- 21: /cgn2_6/prodata/1/pubpna/US60_NEW_PUB.seq:*
- 22: /cgn2_6/prodata/1/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysts of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1235.8	99.7	1239	US-10-032-585-6493	Sequence 6493, App
2	99.4	8.0	2132	US-10-094-749-726	Sequence 726, App
3	96.6	7.8	2320	US-10-305-720-1054	Sequence 1054, App
4	92	7.4	994	US-10-029-386-20763	Sequence 20763, App
5	92	7.4	2873	US-10-755-889-369	Sequence 369, App
6	91.8	7.4	2110	US-10-104-047-1178	Sequence 1178, App
7	91.8	7.4	3078	US-10-104-047-622	Sequence 622, App
8	90.4	7.3	2114	US-10-094-749-692	Sequence 692, App
9	90.2	7.3	2230	US-10-094-749-499	Sequence 499, App
10	89.8	7.2	590	US-10-029-386-13298	Sequence 13298, App
11	89.2	7.2	2221	US-10-108-260A-29	Sequence 29, App1

12	89.2	7.2	3839	US-10-172-118-910	Sequence 910, App
13	89.2	7.2	3839	US-10-342-887-910	Sequence 910, App
14	89	7.2	3502	US-10-098-841-48	Sequence 48, App1
15	88.6	7.2	1466	US-10-029-386-25074	Sequence 25074, App
16	88.4	7.1	1410	US-10-029-386-20532	Sequence 20532, App
17	88.4	7.1	2509	US-10-381-327-11	Sequence 11, App1
18	88.4	7.1	4132	US-10-198-846-13133	Sequence 13133, App
19	87.6	7.1	583	US-10-029-386-13093	Sequence 13093, App
20	87.6	7.1	1757	US-10-203-058B-6	Sequence 6, App1
21	87.2	7.0	831	US-10-264-049-975	Sequence 975, App1
22	86.8	7.0	262	US-10-363-618-66	Sequence 66, App1
23	86.2	7.0	911	US-10-029-386-22636	Sequence 22636, App
24	86.2	7.0	1422	US-10-029-386-20322	Sequence 20322, App
25	86.2	7.0	1725	US-10-029-386-20556	Sequence 20556, App
26	86.2	7.0	1727	US-10-029-386-20367	Sequence 20367, App
27	86.2	7.0	2558	US-09-764-864-1664	Sequence 1664, App
28	86.2	7.0	2558	US-09-764-864-1665	Sequence 1665, App
29	86	6.9	2664	US-10-108-260A-511	Sequence 511, App
30	86	6.9	2690	US-10-723-860-8039	Sequence 8039, App
31	86	6.9	3400	US-10-104-047-1037	Sequence 1037, App
32	85.4	6.9	2505	US-10-108-260A-808	Sequence 808, App
33	84.8	6.8	2239	US-10-094-749-1084	Sequence 1084, App
34	84.4	6.8	1191	US-10-029-386-22992	Sequence 22992, App
35	84.4	6.8	3138	US-10-104-047-83	Sequence 83, App1
36	84.2	6.8	1578	US-10-029-386-20214	Sequence 20214, App
37	84	6.8	565	US-10-029-386-13970	Sequence 13970, App
38	84	6.8	1619	US-10-422-522-45	Sequence 46, App1
39	84	6.8	3617	US-10-029-386-22868	Sequence 22868, App
40	84	6.8	5820	US-10-363-616-44	Sequence 44, App1
41	83.6	6.7	596	US-10-029-386-6664	Sequence 6664, App
42	83.2	6.7	1101	US-10-029-386-25405	Sequence 25405, App
43	83.2	6.7	2241	US-10-108-260A-1289	Sequence 1289, App
44	83.2	6.7	2525	US-10-108-260A-1358	Sequence 1358, App
45	83	6.7	500	US-10-029-386-6818	Sequence 6818, App

ALIGNMENTS

RESULT 1					
US-10-032-585-6493					
; Sequence 6493, Application US/10032585					
; Publication No. US20030180953A1					
GENERAL INFORMATION:					
; APPLICANT: Terry, Koemer D.					
; APPLICANT: Bo, Jiang					
; APPLICANT: Charles, Boone					
; APPLICANT: Howard, Bussey					
TITLE OF INVENTION: Gene Disruption Methodologies for Drug Target Discovery					
FILE REFERENCE: 10102-005-999					
CURRENT APPLICATION NUMBER: US/10/032,585					
CURRENT FILING DATE: 2001-12-20					
NUMBER OF SEQ ID NOS: 8000					
SOFTWARE: PatentIn version 3.1					
SEQ ID NO 6493					
LENGTH: 1239					
TYPE: DNA					
ORGANISM: Candida albicans					
US-10-032-585-6493					
Query Match					
Best Local Similarity 99.8%; Pred. No. 2,6e-227;					
Matches 1237; Conservative 0; Mismatches 2; Indels 0; Gaps 0;					
QY					
1 ATGAGTGAAGTCAGCAACCAATGCATATCTTTAATATCTTCTTTCATCA 60					
Db					
1 ATGAGTGAAGTCAGCAACCAATGCATATCTTTAATATCTTCTTTCATCA 60					
QY					
61 CGTCCCAAAAGTATTTTGACATATGGAAGGTGTGTAAGCCATATATGACCATCA 120					
Db					
61 CGTCCCAAAAGTATTTTGACATATGGAAGGTGTGTAAGCCATATATGACCATCA 120					
QY					
121 TTATTAGACACATTTAAGAACCCACAGTAATGATCGACCGTAAATGTAACAGTGAC 180					

Db 121 TTAATGAGCAATTAAAGAACCCAGTAATGACCGGATTAATGTAACAGGAG 180
QY 181 GATTGATTAAGCATTTTTCAGAAAATCATTTTGGAAAACATATTGTATCATATCC 240
Db 181 GATTGATTAAGCATTTTTCAGAAAATCATTTTGGAAAACATATTGTATCATATCC 240
QY 241 GAAAAAAACCATTCATTTGTCAGTGTGTGTAAAGGGTTAATCTCGACAACCTTG 300
Db 241 GAAAAAAACCATTCATTTGTCAGTGTGTGTAAAGGGTTAATCTCGACAACCTTG 300
QY 301 AAAAGCATGAAATCACCCATACAAAGTCATTTAAATGTCATTTGAAAAATTTGCAAGA 360
Db 301 AAAAGCATGAAATCACCCATACAAAGTCATTTAAATGTCATTTGAAAAATTTGCAAGA 360
QY 361 GCATTTTAAATCATCAATCTTTAAAGCATATATATTATCTGTGATGAAAAAACAATTA 420
Db 361 GCATTTTAAATCATCAATCTTTAAAGCATATATATTATCTGTGATGAAAAAACAATTA 420
QY 421 ACCTGTAAACAATGTATAAAGTTTCACTCGACCTTCAAAATTAAGACACAACTAAATTA 480
Db 421 ACCTGTAAACAATGTATAAAGTTTCACTCGACCTTCAAAATTAAGACACAACTAAATTA 480
QY 481 AAAATCATGATGATGATCTCTGCTTATCAATGTGATCATCTGTTGTTTAAAAATTTTC 540
Db 481 AAAATCATGATGATGATCTCTGCTTATCAATGTGATCATCTGTTGTTTAAAAATTTTC 540
QY 541 CAACTGTGCTATTTTAACTTCAATTAATTCCTCAGATGATGATGATGATGATGATGATG 600
Db 541 CAACTGTGCTATTTTAACTTCAATTAATTCCTCAGATGATGATGATGATGATGATGATG 600
QY 601 AAATGTGTAAAGTGTGTGTGGAAAAAGTTTATCTTCAATATGTTAAAGTCATGAT 660
Db 601 AAATGTGTAAAGTGTGTGTGGAAAAAGTTTATCTTCAATATGTTAAAGTCATGAT 660
QY 661 GATTCTACATGATCAAAATATGACTGTGATTAATGTGATGTGGGAAATTTGCAAG 720
Db 661 GATTCTACATGATCAAAATATGACTGTGATTAATGTGATGTGGGAAATTTGCAAG 720
QY 721 AAAAATGAATTAATTTAGAACATTAATATCTTCAGATGATGATGATGATGATGATGATTA 780
Db 721 AAAAATGAATTAATTTAGAACATTAATATCTTCAGATGATGATGATGATGATGATGATTA 780
QY 781 TTAAGGAAATCTGAAGTGAATAAATTAAGAACCTTATGATCAAGATGAAATTAAT 840
Db 781 TTAAGGAAATCTGAAGTGAATAAATTAAGAACCTTATGATCAAGATGAAATTAAT 840
QY 841 AAATTCATGATTAAGAACAGAAATTAAGATGAGAAAGATGAAAGATGAAAG 900
Db 841 AAATTCATGATTAAGAACAGAAATTAAGATGAGAAAGATGAAAGATGAAAG 900
QY 901 GATATGCTAGATGAAAAAAGATGATGATGATGATGATGATGATGATGATGATGATGATG 960
Db 901 GATATGCTAGATGAAAAAAGATGATGATGATGATGATGATGATGATGATGATGATGATG 960
QY 961 ATTAATCATTTTCTGCTTTTGAAGGTTCAAGAGTGTCTTAACTTATTTCTGAAT 1020
Db 961 ATTAATCATTTTCTGCTTTTGAAGGTTCAAGAGTGTCTTAACTTATTTCTGAAT 1020
QY 1021 AGTGGAAAGATCAATTTGCTTAAGAAATTAATGATGATGATGATGATGATGATGATGAT 1080
Db 1021 AGTGGAAAGATCAATTTGCTTAAGAAATTAATGATGATGATGATGATGATGATGATGAT 1080
QY 1081 GATTTCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1140
Db 1081 GATTTCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1140
QY 1141 AATAGATAGAAAAAGAAATCCAGAGATGAAATGATGATGATGATGATGATGATGATGATGAT 1200
Db 1141 AATAGATAGAAAAAGAAATCCAGAGATGAAATGATGATGATGATGATGATGATGATGATGAT 1200
QY 1201 GATTATTTGCAATGAAATCATGATGATTTCTGATTA 1239
Db 1201 GATTATTTGCAATGAAATCATGATGATTTCTGATTA 1239

Db 1201 GATTATTTGCCAAATGAAAACATCAGTGAATTTCTCGATTA 1239
RESULT 2
US-10-094-749-726
Sequence 726, Application US/10094749
GENERAL INFORMATION:
APPLICANT: ISOGAI, TAKAO
APPLICANT: SUGIYAMA, TOMOYASU
APPLICANT: OTSUKI, TETSUJI
APPLICANT: WAKAMATSU, AI
APPLICANT: SATO, HIROYUKI
APPLICANT: ISHII, SHIZUKO
APPLICANT: YAMAMOTO, JUN-ICHI
APPLICANT: ISONO, YUUKO
APPLICANT: HIO, YURI
APPLICANT: OTSUKA, KAORU
APPLICANT: NAGAI, KEIICHI
APPLICANT: IRIE, RYOTARO
APPLICANT: TAMECHIKA, ICHIRO
APPLICANT: SEKI, NAOHICO
APPLICANT: YOSHIKAWA, TSUTOMU
APPLICANT: OTSUKA, MOTYUKI
APPLICANT: NAGAHARI, KENJI
APPLICANT: MASUHO, YASUHIKO
TITLE OF INVENTION: NOVEL FULL-LENGTH cDNA
FILE REFERENCE: 08435/0160
CURRENT APPLICATION NUMBER: US/10/094,749
CURRENT FILING DATE: 2002-03-12
PRIOR APPLICATION NUMBER: 60/350,435
PRIOR FILING DATE: 2002-01-24
PRIOR APPLICATION NUMBER: JP 2001-328381
PRIOR FILING DATE: 2001-09-14
NUMBER OF SEQ ID NOS: 381
SOFTWARE: Patentin Ver. 2.1
SEQ ID NO 726
LENGTH: 2132
TYPE: DNA
ORGANISM: Homo sapiens
US-10-094-749-726
Query Match 8.0%; Score 99.4; DB 17; Length 2132;
Best Local Similarity 50.6%; Pred. No. 4e-09;
Matches 297; Conservative 0; Mismatches 281; Indels 9; Gaps 2;
QY 7 GAAAGTGACGAACCAATTCATATCATTTTAAATATCTTCTTCTTCTTCAACGCTCC 66
Db 1384 GAATGGGCAAGCTTTTAAACCAAGTCTTCAATCTTACTACACATPAAGAAATTCATACT 1443
QY 67 AAAAGTATATTTGACATATGAGAGGTGTATAAGCCTATATGACCATCATATTA 126
Db 1444 GAGAGAAATCTTACAAATGTAAAGATGTGGCAAGGCTTTCTATCGATCTTCAAACTT 1503
QY 127 GAGCAATTTTAAAGAACCAAGTATGATGATGATGATGATGATGATGATGATGATGATG 186
Db 1504 ACTGAACATPAAGAAATTTATCTGAGAGAAACCTTACATGT-----GAAAGATGT 1557
QY 187 GATPAAGATTTTTCAGAAAAATCAATTTGGAACACATATTTGATACATTTCCGAAAA 246
Db 1558 GCGAAAGCCTTTTAAACCACTCTTCAACCTTGTACACATPAAGTATTCATCTGAGAG 1617
QY 247 AAAACATTCATTTGTCAGTGTGTGTAAAGGGTTAATTTCTGCAACACTGAAAAA 306
Db 1618 AAACCTTCAATGTAAGAAATGTGTAAAGCCTTTAACAGTCTTCAACCTTACTTAA 1677
QY 307 CATGAATCACCATACAAAGTCATTTAATGTATGATTTGAAATTTGCAAGAGCATTT 366
Db 1678 CATPAAGATTCATATCTGAGAGAAAGAAACCTTCAATGTGAAATTTGTGCAAGCTTTT 1737
QY 367 TATPAACATCAATCTTTAA--GACATCATATATTTCTGTTCATGAAAAAACATTAAG 423
Db 1738 AACCACTCTCAACCTTACTGACATPAAGAAATTCATATCTGTGAGAAACCTTACAAA 1797

OY	424	TGTAACATGTAAATAAAGTTTCAGCTGCAGCTTCAAAAATTAGACAACATAATTAAAA	483
Db	1798	CCTAAAGAATGTAAACGAATTTTGAAAAACCTTCAAAGTTTTCTAAACATTAAGAATAAT	1857
OY	484	CATCATGGTGCATCTCGCTTATCAATGTGATCATCTGGTGTGTTTAAAAATTTCCAA	543
Db	1858	TATGCTGGTGAAGAATCTTAGAAATGTGAAGAATGTAAACAAACTTTAAAGTTGCCAC	1917
OY	544	ACTTGTCATGTAATTAACAATTTTCATATTAACAACATCTGCATCCAACACT	590
Db	1918	ACTGATGTGTCATATAGATTAATTCATCTAAAAAACCCTATAGT	1964
RESULT 3			
US-10-305-720-1054			
; Sequence 1054, Application US/10305720			
; Publication No. US20040010136A1			
; GENERAL INFORMATION:			
; APPLICANT: Au-Young, Janice K.; Sellhammer, Jeffrey J.			
; TITLE OF INVENTION: Composition for the Detection of Signaling Pathway Gene Expression			
; FILE REFERENCE: PA-0002-1 CON			
; CURRENT APPLICATION NUMBER: US/10/305,720			
; CURRENT FILING DATE: 2002-11-26			
; PRIOR APPLICATION NUMBER: 09/016,434			
; PRIOR FILING DATE: 1998-01-30			
; NUMBER OF SEQ ID NOS: 1490			
; SOFTWARE: PERL Program			
; SEQ ID NO 1054			
; LENGTH: 2320			
; TYPE: DNA			
; ORGANISM: Homo sapiens			
; FEATURE:			
; NAME/KEY: misc feature			
; OTHER INFORMATION: GenBank ID No. US20040010136A1 g1017721			
US-10-305-720-1054			
Query Match 7.8%; Score 96.6; DB 17; Length 2320;			
Best Local Similarity 51.3%; Pred. No. 1.4e-08;			
Matches 306; Conservative 0; Mismatches 279; Indels 12; Gaps 3;			
OY	69	AAAGTATTTGGCAGATATGAAAGGTGTGATAAAGCCCTAATATGACATCATTAATTAGA	128
Db	964	AGAGAACCCCTACAAATGTAAAGATGTGTAAGCTTTTAAACGATCTTCAACCCCTTAC	1023
OY	129	GCAACATTTAAGAACCCGACAGTAATGATCGACCGTAAATATGTAAGTGACATGTGTA	188
Db	1024	TACCCATGAAAAATTCATCTACTGAGAGAAACCTTCAAAATGT-----GAGATGTGG	1077
OY	189	TAAAGCATTTTTCAGAAAAATCACATTTGGAAACACATATGTATCATATTCGAAAAAAA	248
Db	1078	CAAGCCTTTAGACAGTCTCAAACTTACTACACATAGATTAATCTACTGTGAGAGAA	1137
OY	249	ACCATTCCATTGTTCAAGTGTGTGTAAAGGGTTAATTTCTGACAACACTTGAAGAGACA	308
Db	1138	ACCCATCAAAATGTAAAAAATGTGAAAAAGCCCTTTTACCAAGTCTGACACCTTACACACA	1197
OY	309	TGAATACCCCATTAAGATCAATTTAAATGTATCAATTTGAAAAATGTGAAGAGACTTTA	368
Db	1198	TGAGGTATTTATCTACTGAGAGAAACCTTACAAATGTGAAAAATGTGAAAAAGCTTTAA	1257
OY	369	TAAACA--TCATCTTTAAGACATCATATATATATCTGTTTCATGAAAAAATCATTAACGTG	425
Db	1258	TCATTTTCACACCTTACTACACATTAAGATTAATTCATCTACTGAGAGAAACCTTACAAATG	1317
OY	426	TAAACATGTATTAAGTTTTCACCTCGACCTTCAAAATTTAGACAACAATAATTAAACA	485
Db	1318	TAAAGAAATGTGTAAAGCTTTTAAACACTCTTCAACCTTACTAAACATTAAGATAATTCA	1377
OY	486	TCATGATGATCTCCTGCTTATCAATGTGATCATCTCGTGTGTTTAAAAATTTCCAAC	545
Db	1378	TACTGAGAGAAAGCTTACAAATCTTAAGAAATGTGAAAAAGCTTTTAAACCAATCTTCAA	1437

Oy	546	TTCGCTGCATATTCAATTTCCATATATAAACAAGTGTACATCAGAAACTTAATGTCTCTAATG	605
Dd	1438	ACTTAGACATCATGAAGAAAATTTCACTCGAGAGAAACCCTA---TGAAATGGAAAAATG	1494
Oy	606	TGCTAAAGCTTGTGTGGGAAAAAGGTTATATCTTCAATATGTGAATCATATGATGA	662
Dd	1495	TGGCAAAGCTTTTAAACGAGTCTCCAATCTTACTAGACATTAAGAAGATCATACAGA	1551

RESULT 4
US-10-029-386-20763/c
; Sequence 20763, Application US/10029386
; Publication No. US20030194704A1
GENERAL INFORMATION:
APPLICANT: Penn, Sharon G.
APPLICANT: Rank, David R.
APPLICANT: Hanzel, David K.
TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
FILE REFERENCE: ABOVICA-X-2
CURRENT APPLICATION NUMBER: US/10/029,386
CURRENT FILING DATE: 2001-12-20
NUMBER OF SEQ ID NOS: 34288
SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
SEQ ID NO 20763
LENGTH: 994
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
OTHER INFORMATION: MAP TO AC010620.3
OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 2.4
OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 3.1
OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 3
OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 4.1
OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 2.6
OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 3.4
OTHER INFORMATION: NT HIT: g13650821, EVALUE 0.00e+00
OTHER INFORMATION: EST HUMAN HIT: BI093435.1, EVALUE 0.00e+00
OTHER INFORMATION: SWISSPROT HIT: Q03923, EVALUE 0.00e+00
US-10-029-386-20763

Query Match	7.4%; Score 92;	DB 16;	Length 994;
Best Local Similarity	52.0%; Pred. No. 7.7e-08;		
Matches 259;	Conservative 0;	Mismatches 220;	Indels 9; Gaps 2;

Oy	7	GAAAGTACGAACCAAAATGCATATCAATCTTAAATATCTTCTTCTTCAATCACGCC	66
Dd	763	GAAATGTGGCAAAAGCTTCTACCATCTTTCACACCTTCAACATTAAGSTAAATTCADACT	704
Oy	67	AAAAAGTATATTGACATATGAAGGGTGTAAAGCATATATGACGACATCTTATTA	126
Dd	703	GGAAGAAAGCCCTTCAAAATGTGAAGAAATGTGTAAAGCTTTAACCAACCTTCAGCCCTT	644
Oy	127	GAGCAACATTTAAGAACCCACAGTAAATGATCGACCGTAAATGTAGACGTGACGATTTG	186
Dd	643	ACTACACATTAAGTTCATTCATGTTAAAGAAAAAACCTTACAAATGT-----GAAGAAATGT	590
Oy	187	GATAAAGCATTTTTCAAGAAAATGCATTTGGAAAACATATTTGATCATCATTTCCGAAAA	246
Dd	589	GACAAACCTTTTAAACCAATTTCTCATACCTTACTTAACATTAAGATTAATTCATTTGGAGAG	530
Oy	247	AAACCATTTCAATTTGAGTGTGTGTAAAGGGGTAAATCTTGACAACACTTGAAAGA	306
Dd	529	AAATCTTACAAATGTGAACATGTGGCAAGGCTTTAACTGTGTTTCAACCTTACAAAA	470
Oy	307	CATGAATTCACCCATACAAAGTCATTTAAATGTATCATTTTGAAAATTTGTCAAGAAAGATT	366
Dd	469	CATAAAGAATTCATACGAGAGAAACCCATCAAAATGTGAAGATGTGGCAAGCCTTT	410
Oy	367	TATTAACAT---CAATCTTTAAGCATCATATATATCTGTCTCATGAAAAAACAATTACG	423
Dd	409	AATGTGTCTTCAACCTTACTACCATTAAGATGAATGATCATCTGAGAGAAACCTTACAAA	350

Qy	424	TGAAACAATGTAATAAGTTTCACTGACCTTCAAAATTAGACAACTAAATTAAA	483
Db	349	TGTGAAGAAAGTGGCAAGCCTTTAACCACTCTTAAACTTAATATGATTAATTT	250
Qy	484	CATCATGTGTGATCTCTCT	501
Db	289	CATATCTGGAGAGAAACCT	272

RESULT 5

```

US-10-755-889-369
; Sequence 369, Application US/10755889
; Publication No. US20040171823A1
; GENERAL INFORMATION:
; APPLICANT: Bristol-Myers Squibb Company
; TITLE OF INVENTION: POLYNUCLEOTIDES AND POLYPEPTIDES ASSOCIATED WITH THE NF-KB
; TITLE OF INVENTION: PATHWAY
; FILE REFERENCE: D0284 NP
; CURRENT APPLICATION NUMBER: US/10/755,889
; CURRENT FILING DATE: 2004-01-13
; PRIOR APPLICATION NUMBER: U.S. 60/440,068
; PRIOR FILING DATE: 2003-01-14
; PRIOR APPLICATION NUMBER: U.S. 60/469,757
; PRIOR FILING DATE: 2003-05-12
; NUMBER OF SEQ ID NOS: 823
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 369
; LENGTH: 2873
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-755-889-369

```

Query Match 7.4%; Score 92; DB 18; Length 2873;

Best Local Similarity 52.2%; Pred. No. 1.2e-07;
Matches 256; Conservative 0; Mismatches 225; Indels 9; Gaps 2

QY 7 GAAAGTGAGAAACCAATCGATATCATCTTTAATATCTTCTTCTTATCAGTCCC 66
Db 1621 GAAGTCGCAAAACATTTTGCATGCTTTCACACCTTAACATTAAGAAGATCCAACT 1686

QY 67 AAAAAGTATATTGGACATATGAAGGGGTGATTAAGCCATTAATCGACCATATTATTA 126

Db 1681 AGAGTGAATTTCTACAAATGTGAACATATGAAGAAGCCCTTAATCTGTCCTCAACCTT 1744

0y GAGCAACATTTAAGAACCCACAGTAATGATCGACCGTATTAATGTA CAGTGGACATTTG 186
127
1741 AATAAACATATAGAGAAATTCATCTCGAGAGAAAACTTACAAATGTAAA-----GATGTG 179

Oy	187	GATTAAGCATTTTTCAGAAATCATTTGGAAACATATTTGATCATTTCCGAAAA	246
Db	1795	GGCAAGCCTTTAACGAGCCTCACTTATTAGACATTAAGAGATTCTACTGAAG	185

QY 247 AAACCATTCATGTTGAGGTGTGCTAAAGGGGGTAAATTCTCGCAACACTTGAAAAA 306

Db 1855 AAACCTACAAATGTGAGAGATGTGTCAAAAGCCTTTAACCAAGTCATCGACCTTACTACA 1914

Oy	307	CATGAATAACCCCATCAAAAGTCATTTAATGTACATTTGAAAATTTGCAGAAGCATT	366
Db	1915	CATAATATAATTCATACTGGGAAATTCCTACAATGTAGAAAAATGTTAGACTT	1977

QY 367 TATAAATCAATCTTTAAGACATATATATATATCTGTT---CATGAATAAACAATTAAAG 423

Db 1975 AACCAAGCCTCAAAAGCTTACTGACATTAAGTTAATTCATACCGGAGAGAAACGTTATGAA 203

Qy	424	TGTTAAACAAATGTAAATTAAGTTTTCAGCTCGACCTTCAAAATATAGCAACAAGTAAATTTAAA	483
Db	2035	TGTATGAAGATGCGCAAAAGTTTAAACCGATCTTCAAAACCTTACTGGAACATTAAGTACATT	2094

QY	484	CATCATGGTG	493
Db	2095	CATACTGGAG	2104

RESULT 6
US-10-104-047-1778
; Sequence 1778, Application US/10104047
; Publication No. US20030236592A1

TITLE :

```

, FILE REFERENCE: H1-A0105
, CURRENT APPLICATION NUMBER: US/10/104,047
, CURRENT FILING DATE: 2002-03-25
, PRIOR APPLICATION NUMBER:
, PRIOR FILING DATE:
, NUMBER OF SEQ ID NOS: 4096
, SOFTWARE: PatentIn Ver. 2.1
, SEQ ID NO 1778
, LENGTH: 2110
, TYPE: DNA
, ORGANISM: Homo sapiens
, US-10-104-047-1778

```

Query Match	7.48;	Score 91.8;	DB 17;	Length 2110;
-------------	-------	-------------	--------	--------------

Best Local Similarity 51.7%; Pred. NO. 1.1e-01;
Matches 314; Conservative 0; Mismatches 277; Indels 16; Gaps 4;

QY 61 CGTCCCAAAAAGATATATTTGGACATATGAAGSGTGATTAAGGCTTAATCGACCATCA 120

Db 1253 CATACCGAGAGAAACCTTACCAATGTAAAGATGTGGCAAGCTTTTAAACCTCCTCA 1312

QY 121 TTATTAGGCACTTTAAGAACCCACGTAATGATGCAACCGTAAATGTCAGTGGAC 180
Db 1313 GCCCTTACTACATAGAAGAAATTACACTGGAGAGAAACCCCTCAAAATGT-----GAA 1366

QY 181 GATTGATAAAGCATTTTTCAGAAATTCACATTTGGAAACACATATTGTATCATATCC 240
 Db 1367 GAATGTGCAGAAAGCTTTTAAACGATCCTCAAACCTTACTGAACATATAAGAACTTCATCT 1426

Qy 241 GAAAAAAAAACCAATTCATTGTCAGTGTGGTAAAGGGGTAATTCGACACACTTG 300

Db 1427 GGAAGAAACCTACAAATGTGAAGANTGTGCAGAAAGTTTATTCAAATCCTCAAACTT 1486

QY 301 AAAAGACGTGAATCACCCTACCAAGCATTTAAATGTAACFTTGAATTTGTCAAGAA 360

Db 1487 ACTGAACATTAATAATTCATCTGTGAGAGATACCTTACACAGTGTGAAGATTTGGCANA 1546

QY 361 GCATTTTTAAACATCAATCTTTA---AGACATCATATTTATCTGTCATGAAAAACA 417

Db 1547 GCTTTTAAACATCTCTCATCTCCCTTACTACACATTTAAAGAAATTCATCTACTGGGAGAAACCC 1606

QY 418 TTAACGCTAAACAATGTAATTAAGTTTCACGACCTTCAAAATATAGCACACATATA 477

Db 1607 TACCAATGTGAAGAATGTGGCAAGCTTTTATGCGATCTCAAACTTAATCTGAACATATAG 1666

QY 478 TTTAAACATCATGSGT-GATCCTCGCTATCATCATGATCACTCGTGTGTTTAAAAA 536

Db 1667 ATATTTACTACTGAGAGAAACCTTATTAATGTGAGAGATGTACAAAGCTTTTAAACCA 1726

QY 537 TTTCGAACCTGGTCAGTATTCATTTCACTPAAAAACAAGCATCCAAAACCTTAATG 596
Db 1727 TCTGCAACCT-----TACTAACACATPAAAAATACACTCTCGAGAGAACTACAGAAC 17800

Qy 597 TCTAAATGTGGAAAGGTGCTGTGGAAAAAAGGTTATCTTCACATATGTTAAGCA 656
Db 1781 TGGAAATGTGTGATATGATATTTTGACACACCTCAAAATTTTCTTAAATATTAAGGAAATCA 1840

QY	657	TGATGAT	663
Db	1841	TACTGT	1847

RESULT 7
US-10-104-047-622

RESULT 7

US-10-104-047-622
; Sequence 622, Application US/10104047
; Publication No. US20030236392A1

```

; GENERAL INFORMATION:
; APPLICANT: HELIX RESEARCH INSTITUTE
; TITLE OF INVENTION: No. US20030236392A1el full length cdna
; FILE REFERENCE: HI-A0105
; CURRENT APPLICATION NUMBER: US/10/104,047
; PRIOR FILING DATE: 2002-03-25
; PRIOR APPLICATION NUMBER:
; NUMBER OF SEQ ID NOS: 4096
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 622
; LENGTH: 3078
; TYPE: DNA
; ORGANISM: Homo sapiens
; US-10-094-749-622

```

```

Query Match      7.4%; Score 91.8; DB 17; Length 3078;
Best Local Similarity 52.1%; Pred. No. 1.3e-07;
Matches 257; Conservative 0; Mismatches 227; Indels 9; Gaps 2;

```

```

QY 11 GTGACGAAACCAATGATATCATCTTAATATCTTCTTCTGATCAAGTCCCAAA 70
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 650 GTGGCAATATATTTTGATGATGCTTTCACATTAATCACTCAAGAAATTCATTAAG 709
QY 71 AGTATATTTGACATATGAAGGGTGTGATGAAGCTTAATTCAGCATCATTTAGAGC 130
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 710 AGTATCTTCAATATGGAAGATGTGTAAAGCTTTAAGTGTCTGATCAACCTTAATA 769
QY 131 AATATTTAAGAACCCACAGTAATGATGACCGTAATAATGTACAGTGGACGATTGTGATA 190
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 770 AATATAGATTAATTCATCTACTGAGAGAAACCCCTACAAATGT-----GAAGATGTGGCA 823
QY 191 AAGCATTTTTCAGAAATATCATTTTGGAAACATATTTGATCAATTCGGAAGAAAC 250
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 824 AAGCTTTTAAACCGGTCTCTCAATCTTCTTAACATTAATAATTCATCTGAGAGAAAC 883
QY 251 CATTCATTTGTCAGTGTGTGTAAGGGGTAAATTTCTGACAAACCTTGAAAGACATG 310
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 884 CCAACAAATGTGAAGATGTGGCAAGCTTTTAAACGGTCTGACCCCTTACCTTAACATA 943
QY 311 AATACACCCATACAAAGTCATTTAAATGATCATTTGAAATTTGTCAAGAAACATTTTATA 370
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 944 AAGAAATTCATACAGAAAGAAACCCCTACAAATGTGAAGATGTGGCAAGGCTTTAAAC 1003
QY 371 AATCATATCTT---AAGCATCATATATATCTGTCTGAAAAACCTTAAGCTGA 427
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1004 AGTTCGATCTTTAATTAACATTAAGAAATTCATATGAAATTAACCTTACAAATGTG 1063
QY 428 AACATGTATTAAGTTTTCATCTGACCTTCAAAATTTAGCAACAATTAATTAACATC 487
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1064 AAGAAATGTGGCAAGCTTTTGAAGTATCTCAATTTCTTAATAAACAATTAAGATATCATTA 1123
QY 488 ATGTGATCTCC 500
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1124 CTGGGAGAAAACC 1136

```

```

RESULT 8
US-10-094-749-692
; Sequence 692, Application US/10094749
; Publication No. US20030219741A1
; GENERAL INFORMATION:
; APPLICANT: ISOGAI, TAKAO
; APPLICANT: SUGIYAMA, TOMOYASU
; APPLICANT: OTSUKI, TETSUJI
; APPLICANT: WAKAMATSU, AI
; APPLICANT: SATO, HIROYUKI
; APPLICANT: ISHII, SHIZUKO
; APPLICANT: YAMAMOTO, JUN-ICHI
; APPLICANT: ISONO, YUUKO
; APPLICANT: HIO, YURI
; APPLICANT: OTSUKA, KAORU
; APPLICANT: NAGAI, KEIICHI

```

```

; APPLICANT: IRIE, RYOTARO
; APPLICANT: TAMECHIKI, ICHIRO
; APPLICANT: SEKI, NAOKIHO
; APPLICANT: YOSHIKAWA, TSUTOMU
; APPLICANT: OTSUKA, MOTYUKI
; APPLICANT: NAGAHARI, KENJI
; APPLICANT: MASUHO, YASUHIKO
; TITLE OF INVENTION: NOVEL FULL-LENGTH CDNA
; FILE REFERENCE: 084335/0160
; CURRENT APPLICATION NUMBER: US/10/094,749
; PRIOR FILING DATE: 2002-03-12
; PRIOR APPLICATION NUMBER: 60/350,435
; PRIOR FILING DATE: 2002-01-24
; PRIOR APPLICATION NUMBER: JP 2001-328381
; PRIOR FILING DATE: 2001-09-14
; NUMBER OF SEQ ID NOS: 3381
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 692
; LENGTH: 2114
; TYPE: DNA
; ORGANISM: Homo sapiens
; US-10-094-749-692

```

```

Query Match      7.3%; Score 90.4; DB 17; Length 2114;
Best Local Similarity 51.8%; Pred. No. 2.1e-07;
Matches 258; Conservative 0; Mismatches 231; Indels 9; Gaps 2;

```

```

QY 7 GAAAGTGACGAACCAATGATATCATCTTTAATATCTTCTTCTTCATACGCTCC 66
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1024 GAATGTGGCAAGCTTCTTCAATCTTCTTCAACCTTCAACATTAAGTAATTCATCTACT 1083
QY 67 AAAAGTATATTTGCAATATGAGGGTGTGAATGAAGCCATTAATGACATCATATTA 126
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1084 GAGAGAAAGCCCTTCAATATGTGAAGATGTGTAAGCTTTTAAACCACTTCAAGCCCTT 1143
QY 127 GAGCAATTTTAAAGAACCAAGTATATGACCGTATTAATGTACAGTGAAGATTGT 186
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1144 ACTACACATTAAGTCAATCTTCAATGTTAAAGAAACCTTCAAAATGT-----GAAGATGT 1197
QY 187 GATTAAGCATTTTTCAGAAATTCATATTTGAAACATATTTGATCAATTCGGAATA 246
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1198 GACAAAGCTTTTAAACGATTTCTCATCTTCACTTAACATTAATTAATTCATCTGAGAG 1257
QY 247 AAACATTCATTTGTCAGTGTGTGTAAGGGGTAAATTTCTGACAAACCTTGAAGA 306
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1258 AATCTTAACAAATGTGAACAAATGTGGCAAGGCTTTTAACTGTCTTCAACCTTACAAA 1317
QY 307 CATGAATACACCATACAAAGTCATTTAATGTACATTTGAAATTTGTCAAGAAAGATT 366
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1318 CATGAAGAAATTCATCTGAGAGAAACCTTACAAATGTGAAGATGTGGCAAGCTTT 1377
QY 367 TATTAACAT---CAATCTTAAAGATCATATATATCTGTTCATGAAAAACATTAAGC 423
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1378 AATGTCTTCAACCTTCACTACATAAGATGATTCATATCTGAGAGAAACCTTACAAA 1437
QY 424 TGTAAACATGTATTAATTAAGTTTCACTGACCTTCAAAATTTAGCAACAATTAATTA 483
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1438 TGTGAAGATGTGGCAAGCTTTTAAACCACTCTTCAAACTTAATTAATTAATTAAT 1497
QY 484 CATCATGTGATCTCTCT 501
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1498 CATCTCTGAGAGAAACCT 1515

```

```

RESULT 9
US-10-094-749-499
; Sequence 499, Application US/10094749
; Publication No. US20030219741A1
; GENERAL INFORMATION:
; APPLICANT: ISOGAI, TAKAO
; APPLICANT: SUGIYAMA, TOMOYASU
; APPLICANT: OTSUKI, TETSUJI
; APPLICANT: WAKAMATSU, AI

```

```
/ APPLICANT: SATO, HIROYUKI
/ APPLICANT: ISHII, SHIZUKO
/ APPLICANT: YAMAMOTO, JUN-ICHI
/ APPLICANT: ISONO, YUUKO
/ APPLICANT: HIO, YURI
/ APPLICANT: OTSUKA, KAORU
/ APPLICANT: NAGAI, KEIICHI
/ APPLICANT: IRIE, RYOTARO
/ APPLICANT: TAMECHIKA, ICHIRO
/ APPLICANT: SEKI, NAOHITO
/ APPLICANT: YOSHIKAWA, TSUTOMU
/ APPLICANT: OTSUKA, MOTOTYUKI
/ APPLICANT: NAGAHARI, KENJI
/ APPLICANT: MASUHO, YASUHIKO
/ TITLE OF INVENTION: NOVEL FULL-LENGTH cDNA
/ FILE REFERENCE: 084335/0160
/ CURRENT APPLICATION NUMBER: US/10/094,749
/ PRIOR FILING DATE: 2002-03-12
/ PRIOR APPLICATION NUMBER: 60/350,435
/ PRIOR FILING DATE: 2002-01-24
/ PRIOR APPLICATION NUMBER: JP 2001-328381
/ PRIOR FILING DATE: 2001-09-14
/ NUMBER OF SEQ ID NOS: 3381
/ SOFTWARE: PatentIn Ver. 2.1
/ SEQ ID NO 499
/ LENGTH: 2230
/ TYPE: DNA
/ ORGANISM: Homo sapiens
US-10-094-749-499
```

```
Query Match 7.3%; Score 90.2; DB 17; Length 2230;
Best Local Similarity 50.4%; Pred. No. 2.4e-07;
Matches 332; Conservative 0; Mismatches 313; Indels 14; Gaps 4;
```

```
QY 7 GAAAGGAGAAACCAATCGATATCATCTTTAATCTCTTCTTCATACGCTCC 66
DB 656 GAATGAGCAAAATCACTTGGCATGCTTTCACGCTTCACTCAATAAAAATTCATCT 715
QY 67 AAAAGTATATTTGACATATGAGGGGTGATTAAGCCATTAATGACCATATTA 126
DB 716 AAGAGAAATTTCTCAAAATGTGAAGGTGTGAAAAACCTTTACTGTCTCAAACTT 775
QY 127 GAGCAACATTTAAGAACCAAGTATGATCGACCGTATTAATGTAAGTGCATG 186
DB 776 TCTAACTTAAGAAATTCATCTGAGAAAAACCTTACAAATGTGAAGTA-----TGT 829
QY 187 GATTAAGCATTTTTCAGAAAATCACTTTGAAAACATATGTAATTCATCCGAAA 246
DB 830 GGAAGACCTTTCAACCAATCTCAATCTTAACTTAACATAGTAATTCGTACTGAGAA 889
QY 247 AAACATTCATGTTCAAGTGTGTGTAAGGGGTTAATTCGACAACACTTGAAAGA 306
DB 890 AAACCTTAATTAATGTGCACTGTGCAAAAGCCTTTAAACAGTCTTCAACCTTACTA 949
QY 307 CATGAATCAACCATCAAAAGTCATTTAATGTACATTTGAAAATTTGCAAGAGCAT 366
DB 950 CATTAATTAATTCATCTGAGAGAAACCTTCAAAATGTGACAAATGTGGCAAGCTT 1009
QY 367 TATTAACATCATC---TTTAAGACATCATATTAATCTGTTCATGAAAAAACATTA 423
DB 1010 AAGAGTCCCAACCTTCAACATCAATGATTAATTAATCTGAGAGAAACCATCAAA 1069
QY 424 TGTAAACATATTAATTAAGTTTCACTGACCTTCAAAATTAAGCAACATTAATTA 483
DB 1070 TGTGAGAAATGTGCAAGCTTTTAACATTAATTAATTAATTAAGCAATTAAGAA 1129
QY 484 CATCATGTGTG-GATCTCTGCTTATCATATGTGATCATCTGCTGTGTTTAAATTTCA 542
DB 1130 TACACTAGAGAAAGCTTCAAAATGTGAAGATGTGGAAAGCCTTTTAAACAGTTTCA 1189
QY 543 AACTGTGTCAGTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 602
DB 1190 ACCCTTATTAACATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1245
```

```
QY 603 ATGTGTAAGGTTGTGGGAAAAAGCTTATCTCATATGTTAAGTCATGATG 661
DB 1246 ATGTGGCAGAGCTTTTAACCAAGTCCGCAAGCTCATGAATTAATTAATCACTG 1304
```

RESULT 10

```
US-10-029-386-13298
/ Sequence 13298, Application US/10029386
/ Publication No. US20030194704A1
/ GENERAL INFORMATION:
```

```
/ APPLICANT: Penn, Sharon G.
/ APPLICANT: Rank, David R.
/ APPLICANT: Hanzel, David K.
/ TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR G
/ FILE REFERENCE: AEWICHA-X-2
/ CURRENT APPLICATION NUMBER: US/10/029,386
```

```
/ PRIOR FILING DATE: 2001-12-20
/ NUMBER OF SEQ ID NOS: 34288
/ SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
/ SEQ ID NO 13298
```

```
/ LENGTH: 590
```

```
/ TYPE: DNA
```

```
/ ORGANISM: Homo sapiens
```

```
/ FEATURE:
```

```
/ OTHER INFORMATION: MAP TO CHR19.3
```

```
/ OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 4
```

```
/ OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 4.5
```

```
/ OTHER INFORMATION: NT HIT: 91475315, EVALUR 0.00e+00
```

```
/ OTHER INFORMATION: SWISSPROT HIT: Q05481, EVALUR 1.00e-119
```

```
/ OTHER INFORMATION: EST_HUMAN HIT: BG619358.1, EVALUR 0.00e+00
```

```
US-10-029-386-13298
```

```
Query Match 7.2%; Score 89.8; DB 16; Length 590;
Best Local Similarity 53.8%; Pred. No. 1.7e-07;
Matches 234; Conservative 0; Mismatches 192; Indels 9; Gaps 2;
```

```
QY 62 GTCCCAAAAGTATTTGACATATGAGGGGTGATTAAGCCATTAATGACCATCAT 121
DB 126 GTGCTAAAGAGAAATCTCAAGTGTGAAGATGTGGCAAACTTTCTATGTCTTCA 185
QY 122 TATTAGAGCAATTTAAGAACCAAGTATGATCGACCGTATTAATGTAAGTGCAG 181
DB 186 CCTTAAGTAAAGATTAAGAGATACACTGAGAGAAACCTTCAAAATGT-----GAAG 239
QY 182 ATGTGATTAAGCATTTTTCAGAAAATCACTTTGAAAACATATTTGATCACTTCC 241
DB 240 AATGTGGCAAGCTTTTGAAGCAATCTTCAACCTTGTAAACATTAAGAAATTCATCTG 299
QY 242 AAAGAAACCATTCATGTTCAAGTGTGTGTAAGGGGTTAATTCGACAACACTTGA 301
DB 300 GAGAGAAACCTTCAAAATGTGAAGATGTGGCAAAAGCTTTAGCCGTTCTTCAACCTTG 359
QY 302 AAAGCATGAATCAACCATCAAAAGTCATTTAATGTACATTTGAAAATTTGCAAGAG 361
DB 360 CTAACATTAAGATTAATCTACTGAGAGAAACCTTCAAAATGTGAAGATGTGGCAAG 419
QY 362 CATTTTATTAATCATATCTTTAAGACATCATATTAATCTGTTCATGA---AAAAAT 418
DB 420 CTTTATGAATTCCTCAACCTTGTGATTAATTAATTAATTAATTAATTAATTAATTA 479
QY 419 TAACTGTAAACATTAATTAAGTTTCACTGCACTTCAAAATTAAGCAACATTAAT 478
DB 480 ACAAATGTAAAGAAATGTGAAGAGCTTTTAAGCGACTCTCAACCTTCAATTAATTA 539
QY 479 TAAACATCATGTGTG 493
DB 540 TATATCATGCTGGAG 554
```

RESULT 11

```
US-10-108-260A-29
```

```
/ Sequence 29, Application US/10108260A
/ Publication No. US20040005560A1
/ GENERAL INFORMATION:
/ APPLICANT: HELIX RESEARCH INSTITUTE
/ TITLE OF INVENTION: No. US20040005560A1el full length cDNA
/ FILE REFERENCE: H1-A0106
/ CURRENT APPLICATION NUMBER: US/10/108,260A
/ CURRENT FILING DATE: 2002-03-27
/ NUMBER OF SEQ ID NOS: 5458
/ SOFTWARE: PatentIn Ver. 2.1
/ SEQ ID NO 29
/ LENGTH: 2221
/ TYPE: DNA
/ ORGANISM: Homo sapiens
US-10-108-260A-29

Query Match      7.2%; Score 89.2; DB 17; Length 2221;
Best Local Similarity 53.4%; Pred. No. 3.7e-07;
Matches 237; Conservative 0; Mismatches 198; Indels 9; Gaps 2;

QY 61 CGTCCCAAAAGATATTTTGACATATGAGGGGTGATTAAGCCTATATGACATCA 120
DB 654 CATACCGAGAGAAACCTTTACATGTAAAGAAATGGCAAGCCTTTAATAGAACTCA 713
QY 121 TTAATAGAGCAATTTAAGACCCAGTAATGATCGACCGTATTAATGACAGTGAG 180
DB 714 TACGTTACTCGATCGAGAAATGCACTACGAGAGAAACCTTTCAAGTGT-----AG 767
QY 181 GATTGTGATTAAGCATTTTTCAGAAATCACTTTGGAAACATATTTGATCATTCC 240
DB 768 GAATGGGCAAAAGCTTTTAAGAGGCTTCACCTTAATCAATGAGAAATCCATAC 827
QY 241 GAAAAAAACCATTCCTGTTGATGTGTGTGTAAGGGGTTAATCTGCACACCTTG 300
DB 828 GGAGAGAAACCTTCAATGTAAAGATGTGGCAAGCCCTTTAAACGGGGCTCACACCT 887
QY 301 AAAAGCATGAATTCACCCATACAAAGTCAATTAATGTCACTTTGAAATTTGCAAGA 360
DB 888 ACTCGCATCAAGAAATCCACTGAGAGAGAGCTCTTCAAAATGTGAATGTGGCAAA 947
QY 361 GCATTT---TATAACATCATCTTTAAGACATCATATTAATCTGTTCAAGAAAAACA 417
DB 948 GCTTTATCTGGGGCTCACACCTTACACATCAGAGAGCTCCATCTGAGAGAAATTC 1007
QY 418 TTAACGTGAACATGTAATAAAGTTTCACTCGACCTTCAAAATTAAGCAACATAA 477
DB 1008 TTCAAAATGTAAAGATGTGGCAAAAGCTTTTACAGAGAGCTCACACCTTACTCAACATCAG 1067
QY 478 TTAACATCATGTGTGATCTCCT 501
DB 1068 AGAATTCATCTACTGAGAGAAACCT 1091

RESULT 12
US-10-172-118-910
/ Sequence 910, Application US/10172118
/ Publication No. US2003022437A1
/ GENERAL INFORMATION:
/ APPLICANT: Dai, Hongyue
/ APPLICANT: He, Yudong
/ APPLICANT: Linsley, Peter
/ APPLICANT: Mao, Mao
/ APPLICANT: Roberts, Chris
/ APPLICANT: Van 't Veer, Laura
/ APPLICANT: Van de Vijver, Marc
/ APPLICANT: Bernards, Rene
/ TITLE OF INVENTION: Diagnosis and Prognosis of Breast Cancer Patients
/ FILE REFERENCE: 9301-175-999
/ CURRENT APPLICATION NUMBER: US/10/172,118
/ CURRENT FILING DATE: 2002-06-14
/ PRIOR APPLICATION NUMBER: 60/380,770
/ PRIOR FILING DATE: 2002-05-14
/ NUMBER OF SEQ ID NOS: 2699
/ TYPE: DNA
```

```
/ SEQ ID NO 910
/ LENGTH: 3839
/ TYPE: DNA
/ ORGANISM: Homo sapiens
/ PUBLICATION INFORMATION:
/ DATABASE ACCESSION NUMBER: NM_003430
/ DATABASE ENTRY DATE: 2001-06-18
US-10-172-118-910

Query Match      7.2%; Score 89.2; DB 17; Length 3839;
Best Local Similarity 53.7%; Pred. No. 4.6e-07;
Matches 234; Conservative 0; Mismatches 193; Indels 9; Gaps 2;

QY 69 AAAGTATATTTGACATATGAGGGGTGATTAAGCCTTAATTCGACATTAATTAGA 128
DB 964 AGAGAAACCTTACAAATGTGAAGATGTGGCAAGCTTTAGCCATTTCTCAACCTTGC 1023
QY 129 GCAACATTTAAGAACCCACAGTAATGATGACCGTATTAATGTACAGTGAAGATTGCA 188
DB 1024 TAAACATAAGAAATTCATATGAGAGAAACCTTCAAAATGT-----GAAGATGTGG 1077
QY 189 TAAAGATTTTTCAGAAATTCATTTGGAACACATATTTGATCATTTCCGAAAAAAA 248
DB 1078 CAAAGCTTTTACGCAATTTCTTACGCCCTTCTAAACATTAAGAAATTCATATCTGAGAGAA 1137
QY 249 ACCATTCCATTTGTCAGTGTGTGTAAGGGGTTAATTCGACACACTTGAAGAGCA 308
DB 1138 ACCCTACAAATGTAAAGAAATGTGGCAAGCTTTTACAAATTCCTCAACCTTGTATCA 1197
QY 309 TGAATACACCATACAAAGTCAATTTAATGTACATTTGAAAATTTGCAAGAACATTTTA 368
DB 1198 TAAGATTAATCATATCGAAGAGAAACCTTCAAAATGTAAAGATGTGACAAACCTTTTA 1257
QY 369 TAAACATCAATC---TTTAACATCATATATTTATTTGTTGATGAAAAAACAATTACGTG 425
DB 1258 GCGACTCTCAACCTTACTTAAACATTAATATACATGCTGAGAGAAACCTTACAAATG 1317
QY 426 TAAACATTAATTAAGTTTCACTGACCTTCAAAATTAAGCAACATTAATTAAGCA 485
DB 1318 TGAAGAAATGTGGCAAGCTTTTATGATGATCTTCAAAATCTTACTATACATTAATTTCA 1377
QY 486 TCATGTGTGATCTCCT 501
DB 1378 TACTGAGAGAAACCT 1393

RESULT 13
US-10-342-887-910
/ Sequence 910, Application US/10342887
/ Publication No. US20040058340A1
/ GENERAL INFORMATION:
/ APPLICANT: Dai, Hongyue
/ APPLICANT: He, Yudong
/ APPLICANT: Linsley, Peter S.
/ APPLICANT: Mao, Mao
/ APPLICANT: Roberts, Christopher J.
/ APPLICANT: Van 't Veer, Laura Johanna
/ APPLICANT: Van de Vijver, Marc J.
/ APPLICANT: Bernards, Rene
/ TITLE OF INVENTION: Diagnosis and Prognosis of Breast Cancer Patients
/ FILE REFERENCE: 9301-188-999
/ CURRENT APPLICATION NUMBER: US/10/342,887
/ CURRENT FILING DATE: 2003-01-15
/ PRIOR APPLICATION NUMBER: 60/298,918
/ PRIOR FILING DATE: 2001-06-18
/ PRIOR APPLICATION NUMBER: 60/380,710
/ PRIOR FILING DATE: 2002-05-14
/ PRIOR APPLICATION NUMBER: 10/172,118
/ PRIOR FILING DATE: 2002-06-14
/ NUMBER OF SEQ ID NOS: 2699
/ SEQ ID NO 910
/ LENGTH: 3839
/ TYPE: DNA
```

```
/ ORGANISM: Homo sapiens
US-10-342-887-910

Query Match      7.2%; Score 89.2; DB 17; Length 3839;
Best Local Similarity 53.7%; Pred. No. 4.6e-07;
Matches 234; Conservative 0; Mismatches 193; Indels 9; Gaps 2;

QY 69 AAAGTATATTTGACATATGAGGGGTGTGATTAAGGCTATATGACCATCATTTATTA 128
DB 964 AGGAAACCCCTCAAAATGTGAAGAAATGTGGCAAAAGCTTTAGCCATTCTCAACCTTGC 1023

QY 129 GCAACATTTAAGAACCCACAGTAAATGATGACCGTATTAATGTACAGTGGACATTTGCA 188
DB 1024 TAAACATTAAGAGATTCATCTGGAGAGAAACCTTCAAAATGT-----GAAGAAATGTG 1077

QY 189 TAAAGCATTTTTCAGAAAATCACTTTGGAAAACACATTTGTATCATTTCCGAAAAAAA 248
DB 1078 CAAAGCTTTTTCAGCATCTTTCAGCCCTTGCTAAACATTAAGAAATTCATCTGGAGAG 1137

QY 249 ACCATTCCATTGTCAGTGTGTGTAAAGGGGTTAAATTCGACAAACACTTGAAGAGACA 308
DB 1138 ACCCTTCAAAATGTAAAGAAATGTGGCAAAAGCTTTTACAAATTCCTCAACCTTGCTAATCA 1197

QY 309 TGAATACCCACATACAAATGATTTAAATGTACATTTGAAAATGTGCAGAAAGCATTTTA 368
DB 1198 TAAAGTAACTCACTACATGAAAGAGAAACCTTCAAAATGTAAAGATGTGACAAACCTTTTAA 1257

QY 369 TAAACATCATC---TTTAAACATCATATATATCTGTTCATGAAAAAACAATTAACGTG 425
DB 1258 GGGACCTCGAACCCCTTACCTAAACATTAATATATCATCTGGAGAGAAACCTTACAAAG 1317

QY 426 TAAACATGTAATTAAGTTTTCATCTGACCTTCAAAATTAAGCAACATTAATTAACA 485
DB 1318 TGAAGATGTGGCAAGCTTTTATGATGATCTTCAATCTTACTATACATTAAGTTTATCA 1377

QY 486 TCATGTGATCTCCT 501
DB 1378 TACTGGAGAAACCT 1393

RESULT 14
US-10-098-841-48
/ Sequence 48, Application US/10098841
/ Publication No. US20020197679A1
/ GENERAL INFORMATION:
/ APPLICANT: Tang, Y. Tom
/ APPLICANT: Liu, Chenghua
/ APPLICANT: Asundi, Vinod
/ APPLICANT: Xu, Chongjun
/ APPLICANT: Zhou, Ping
/ APPLICANT: Ma, Yungqing
/ APPLICANT: Wang, Jian-Rui
/ APPLICANT: Zhao, Qing A.
/ APPLICANT: Ren, Feiyun
/ APPLICANT: Chen, Rui-hong
/ APPLICANT: Wang, Dunrui
/ APPLICANT: Wang, Zhiwei
/ APPLICANT: Wehrman, Tom
/ APPLICANT: Zhang, Jie
/ APPLICANT: Qian, Xiaohong B.
/ APPLICANT: Drmanac, Radoje T.
/ TITLE OF INVENTION: No. US20020197679A1e1 Nucleic Acids and
/ FILE REFERENCE: 784CIP2
/ CURRENT APPLICATION NUMBER: US/10/098,841
/ CURRENT FILING DATE: 2002-03-13
/ PRIOR APPLICATION NUMBER: 09/598,042
/ PRIOR FILING DATE: 2000-06-20
/ PRIOR APPLICATION NUMBER: 09/552,317
/ PRIOR FILING DATE: 2000-04-25
/ PRIOR APPLICATION NUMBER: 09/488,725
/ PRIOR FILING DATE: 2000-01-21
/ NUMBER OF SEQ ID NOS: 331
```

```
/ SOFTWARE: pt_FL_genes Version 1.0
/ SEQ ID NO 48
/ LENGTH: 3502
/ TYPE: DNA
/ ORGANISM: Homo sapiens
/ FEATURE:
/ NAME/KEY: CDS
/ LOCATION: (154)..(2583)
/ NAME/KEY: misc_feature
/ LOCATION: (1)..(3502)
/ OTHER INFORMATION: n = a,t,c or g
US-10-098-841-48

Query Match      7.2%; Score 89; DB 13; Length 3502;
Best Local Similarity 51.7%; Pred. No. 4.8e-07;
Matches 256; Conservative 0; Mismatches 230; Indels 9; Gaps 2;

QY 79 TGCATATGAAGGGTGTGATTAACCTATATGACCATCTATTATGAGCAACATTTA 138
DB 2101 TACAAATGTGAAGATGTGGCAAAAGCCTTTAAGTGTCTCAACCTTATCAATTAAG 2160

QY 139 AGAACCCACAGTATGATGACCCGTATTAATGTACAGTGAAGATTGTATTAAGCATTT 198
DB 2161 ATAATTCTATCTGGAGAGAAACCTTCAAAATGT-----GAAGATGTGGCAAGCTTT 2214

QY 199 TTCAGAAATCACATTTGGAAAACATATTTGTATCATCTTCGAAAAAAACCATTCAT 258
DB 2215 AAAGTGCCTCAACCTTTTACATATTAATATTAATCTGAGAGAGAAACCTTACAAA 2274

QY 259 TGTTCAGTGTGTGTAAAGGGGTTAATCTCGACAACCTTGAAGAAAGCATGAATCAC 318
DB 2275 TGTGAAAAATGTGGAAAGCTTTTAAACGACCTTCAACCTTATTTGAACATTAAGAAAT 2334

QY 319 CATCAAAAGTCAATTAATGTACATTTGAAAATGTCAAGAACATTT---TATAACAT 375
DB 2335 CATCTGAGAGAACCTTCAAAATGTGAAGATGTGCAGAAAGCTTTTACTATTCCTCA 2394

QY 376 CAATCTTAAAGCATCATATATATATCTGTTCATGAAAAAACAATTAACGTATTAACATGT 435
DB 2395 CACCTTAATACATTAAGAGATTCATCTAAAGAGCAACCTTCAAAATGTAAAGATGT 2454

QY 436 AATAAGTTTTCATCTCGACCTTCAAAATTAAGCAACATTAATTAACATCATGTGCA 495
DB 2455 GGCAGAGCTTTCACCAATATTCAAACCTTACACATTAACAAATTCATCTGAGAG 2514

QY 496 TCTCTGTCTTATCATATGATCATCGTGTGTTTAAATTTCCAAACTGTGCAGTA 555
DB 2515 AAATCTTCAAAACCTGAAAGATGTGACATGATTTTGAACACACCTCAACCTTTTCAAC 2574

QY 556 TTACATTTTCAATATA 570
DB 2575 ATTAATTAATTAATA 2589

RESULT 15
US-10-029-386-25074
/ Sequence 25074, Application US/10029386
/ Publication No. US20030194704A1
/ GENERAL INFORMATION:
/ APPLICANT: Penn, Sharon G.
/ APPLICANT: Rank, David K.
/ APPLICANT: Hanzel, David K.
/ TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR G
/ FILE REFERENCE: ABOMICA-X-2
/ CURRENT APPLICATION NUMBER: US/10/029,386
/ CURRENT FILING DATE: 2001-12-20
/ NUMBER OF SEQ ID NOS: 34288
/ SOFTWARE: Anomax Sequence Listing Engine vers. 1.1
/ SEQ ID NO 25074
/ LENGTH: 1466
/ TYPE: DNA
/ ORGANISM: Homo sapiens
```

```
; FEATURE:
; OTHER INFORMATION: MAP TO AC011467.3
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 0.67
; OTHER INFORMATION: NT HIT: g115309152, EVALUE 0.00e+00
; OTHER INFORMATION: SWISSPROT HIT: P35789, EVALUE 0.00e+00
; OTHER INFORMATION: EST_HUMAN HIT: AUI43734.1, EVALUE 0.00e+00
US-10-029-386-25074
```

```
Query Match      7.2%; Score 88.6; DB 16; Length 1466;
Best Local Similarity 53.3%; Pred. No. 4.1e-07;
Matches 237; Conservative 0; Mismatches 199; Indels 9; Gaps 2;
```

```
QY 60 ACCTCCCAAAAGTATATTTGCACATATGAGGGGTGATTAAGCCTATATCGACCATC 119
    |||||
Db 272 ACATACCTGGAAGAAATCTTCAATGTAAAGATGGAAGATGCAATTTGCATGCTTTC 331
    |||||
QY 120 ATTATTAGACCAATTTAAGAACCCACAGTAATGATCGACCGTATTAATGTACAGTGA 179
    |||||
Db 332 ACCTTAGCTCAACATTAAGAAATTCATAGTGAAGAAACCTACAAATGTAAA----- 386
    |||||
QY 180 CGATTGATTAAGCAATTTTTCAGAAATACAAATTTGGAACACATATTGTATCATTC 239
    |||||
Db 387 -GATGTGGGAAGCCTATATATGAGACCTCAAACTTTCTACATTAAGAAATTCATAC 445
    |||||
QY 240 CGAAAAAAACCAATTCATGTTCAAGTGTGTAAAGGGTTAATTCTGACACACTT 299
    |||||
Db 446 TGGAAAGAAACCTTACAAATGCGAAGAGTGTGAAAAAGCCTTTAACGGCTCAGACCT 505
    |||||
QY 300 GAAAAAGCATGAATCACCACATCAAAAGTCATTTAAATGTACATTTGAAAAATTGTCAGA 359
    |||||
Db 506 TACTACACATTAAGATATTCATCTACTGGAAGAAACCTCAAAATGTGAGAGTGGCAA 565
    |||||
QY 360 AGCATTTTAAACAT---CAATCTTAAGACATCAATATATCTGTTGATGAAAAAC 416
    |||||
Db 566 AGCTTTTAACCAATCTGCAAACTTACTACACATTAAGAAATTCATACCTGAGAGAAACC 625
    |||||
QY 417 ATTACGTGAACCAATGTAAATTAAGTTTCACTCGACCTTCAAAATTAAGACACATA 476
    |||||
Db 626 CTACAAATGTGAAGAGTGGCAGAGCTTTTAGCCAGTCTCAACCTTTACTGCACTAA 685
    |||||
QY 477 ATTAAAACATCATGTGTGATCTCCT 501
    |||||
Db 686 GATTAATTCATGTGAGAGAAACCT 710
    |||||
```

Search completed: February 9, 2005, 11:50:32
Job time : 2162.11 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Comugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 9, 2005, 01:45:39 / Search time 3815.62 Seconds
(without alignments)
12360.150 Million cell updates/sec

Title: US-09-831-804-2
Perfect score: 1239
Sequence: 1 atgagtgtaagacgacgaac.....catcagtgattcttcgataa 1239

Scoring table: IDENTITY NUC
Gapop 10.0, Gapext 1.0

Searched: 34239544 seqs, 19032134700 residues

Total number of hits satisfying chosen parameters: 68479088

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST:*
1: gb_est1:*
2: gb_est2:*
3: gb_hic:*
4: gb_est3:*
5: gb_est4:*
6: gb_est5:*
7: gb_est6:*
8: gb_gsa1:*
9: gb_gsa2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	182.2	14.7	628	BZ298000	CG3957.f1
C 2	167.6	13.5	959	8	BZ298000
3	103.8	8.4	801	7	CN759437
4	103.8	8.4	898	7	CN761003
5	98.2	7.9	1609	3	AK032220
6	97.8	7.9	2647	7	BC047646
7	97.6	7.9	745	1	AU123448
8	93.4	7.5	727	7	CO434402
9	92.4	7.5	631	7	CN352904
10	92.4	7.5	2672	3	BC036394
11	92	7.4	591	8	AQ005136
12	92	7.4	687	7	CN362485
13	92	7.4	752	7	CN417582
14	91	7.3	593	8	B99387
15	91	7.3	820	5	BX414627
16	90.8	7.3	666	7	CN297781
17	90.4	7.3	2694	3	BC025227
18	90.2	7.3	2698	3	BC037782
19	90.2	7.3	2253	3	BC028252
20	89.8	7.2	550	5	BP292533
21	89.2	7.2	582	5	BP234906
22	89.2	7.2	604	2	BE161630
23	89.2	7.2	638	7	CN341035
24	89.2	7.2	673	7	CN353930

25	89.2	7.2	3609	3	BC032590
26	89	7.2	2856	3	HSN803730
27	88.6	7.2	535	1	AL705393
28	88.6	7.2	612	5	BP238960
29	88.6	7.2	2174	3	AK033001
30	88.6	7.2	2330	3	BC037426
31	88.6	7.2	2622	3	AK033958
32	88.4	7.1	998	5	BX456765
33	88	7.1	895	5	BU195416
34	87.8	7.1	608	7	CN266758
35	87.8	7.1	4251	3	CR749856
36	87.6	7.1	744	5	BX437291
37	87.2	7.0	579	8	AQ437265
C 38	87.2	7.0	687	8	AQ389266
C 39	87.2	7.0	907	5	BQ423752
40	87	7.0	660	6	CD770140
41	87	7.0	842	5	BU507408
C 42	86.8	7.0	521	8	AQ475626
43	86.8	7.0	571	7	CN280110
44	86.8	7.0	705	8	AQ194282
C 45	86.8	7.0	860	8	AQ749175

ALIGNMENTS

RESULT 1
BZ298000/c 628 bp DNA linear GSS 31-OCT-2002
LOCUS BZ298000
DEFINITION CG3957.f1 Candida glabrata Random Genomic Library
ACCESSION BZ298000
VERSION BZ298000.1 GI:24440936
KEYWORDS GSS.
SOURCE Candida glabrata
ORGANISM Candida glabrata
Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
Saccharomycetales; mitosporic Saccharomycetales; Candida.
REFERENCE 1 (bases 1 to 628)
Wong,S., Fares,M.A., Zimmermann,W., Butler,G. and Wolfe,K.H.
Evidence from comparative genomics for a complete sexual cycle in
the 'asexual' pathogenic yeast Candida glabrata
JOURNAL Genome Biol. 4 (2), R10 (2003)
MEDLINE 22508158
PUBMED 12620120
COMMENT Contact: Wong S
Department of Genetics, Smurfit Institute
Trinity College Dublin
Dublin 2, Ireland
Tel: 353 1 6082319
Fax: 353 1 6798558
Email: swong@tcd.ie
Classes: plasmid ends.
Location/Qualifiers
FEATURES
source
1..628
/organism="Candida glabrata"
/mol_type="genomic DNA"
/strain="CBS 138"
/db_xref="taxon:5478"
/clone="CG3957"
/clone_lib="Candida glabrata Random Genomic Library"
ORIGIN
Query Match 14.7%; Score 182.2; DB 8; Length 628;
Best Local Similarity 59.5%; Pred.No. 7.5e-26;
Matches 348; Conservative 0; Mismatches 228; Indels 9; Gaps 2;
QY 84 ATATGAAGGTGTGTAAGCCTATATGACATCATTTATAGACACATTATAGAC 143
DB 625 ATACGTAATTGTGACAGCCCTTACAGCGCTCTGCTCACAGAACACAGATTC 566
QY 144 CCACAGTATGATCGACCGCTATTAATGTACAGTGCAGATTGTGATTAAGCATTTTCAG 203

```

Db      565 ---AGTGCATCTTGGAGAGAAACCTTGGAAATGTAATCAATGTGAAGTTCAATTACTAA 509
Qy      204 AAAATCACATTTGAAAACACATATTTGATACATTCGGAAGAAAAACCATTCATTTGTC 263
Db      508 AAAAGATCCACTTAGAGAGACCTTATACACACAGATGAAGACCGTTTATTTGTC 449
Qy      264 AGTGTGTGTAAAGGGGTTAATTTCTGACACACATTTGAAAAGACATGAATACTACCATAC 323
Db      448 ATTGTGTGGAAAGGGGCTTATTACTAGGCAACACATGAACACATAGAGTAATCTACAC 389
Qy      324 AAAGTCATTTAAATGACATTTGAAAATTTGCAAGAAAGATTTTAAATCAATCAATCTT 383
Db      388 CAATTCCTTCAATTTGAAATATGAAAGTTGTAATGAAGTTCTTACAGACCTTCATTT 329
Qy      384 AAGACATCATATATTTATCTGTCATGAAAAACATTAAGCTGTAAACATGTAATTAAGT 443
Db      328 AAGGGGTCAATTTTATGAGTTCAATTTACAAAGTCAAAATGTCAGATGACAAAG 269
Qy      444 TTTCACTGACCTTCAAAATTAAGACACATTAATTAATAACATCAT-----GTTGATC 497
Db      268 TTTTCAAAGACCTTACAGACTCAAAATCACAATAGCAACACATTAATCCAGATGTTGT 209
Qy      498 TCCTGCTTATCATGATGATCATCCGTTGTTTAAATTTCCAAATCTGTGCACTAT 557
Db      208 CAATGCTTATCAATGATGATCTTCAATGTTGTTTAAAGTTTCAAAACATGCTGCTGT 149
Qy      558 ACAATTTCAATTAATAACAATGATCAATCAAACTTAATGCTTAATGTTGTAAGTTG 617
Db      148 AAGATTACATGTTAAATGATCATCCGAAGTTAAATGCTTATATGCAAGTAACCTTG 89
Qy      618 TTTTGGGAAAAAAGTTTATCTTCACATATGTTAAGTCATGATGA 662
Db      88 TGTGGGGGAAGACGGTTTAAATATGACATGAATAATCCACGATGA 44

RESULT 2
CN506MMP/c 959 bp DNA linear GSS 17-JUN-2001
LOCUS T7 end of clone AU00A005F10 of library AU00A from strain CBS 3082
DEFINITION of Saccharomyces kluyveri, genomic survey sequence.
ACCESSION AL405671
VERSION AL405671.1 GI:12168715
KEYWORDS GSS.
SOURCE Saccharomyces kluyveri
ORGANISM Saccharomyces kluyveri
Bukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
Saccharomycetales; Saccharomycetaceae; Saccharomyces.
REFERENCE 1 (bases 1 to 959)
AUTHORS Soucie,J.,U., Aigle,M., Artiguenave,F., Blandin,G.,
Bolotin-Fukuhara,M., Bon,E., Brothier,P., Casaregola,S.,
de-Montigny,J., Dujon,B., Durrens,P., Lepingle,A., Llorente,B.,
Malbertuy,A., Neuvéglise,C., Olier-Kalogeropoulos,O., Porter,S.,
Saurin,W., Tekala,F., Toffano-Nicche,C., Wesolowski-Louvel,M.,
Wincker,P. and Weissenbach,J.
TITLE Genomic exploration of the hemiascomycetous yeasts: 1. A set of
JOURNAL FEBS Lett. 487 (1), 3-12 (2000)
MEDLINE 20584711
PUBMED 11152876
REFERENCE 2 (bases 1 to 959)
AUTHORS Neuvéglise,C., Bon,E., Lepingle,A., Wincker,P., Artiguenave,F.,
Galliardin,C. and Casaregola,S.
TITLE Genomic exploration of the hemiascomycetous yeasts: 9.
JOURNAL FEBS Lett. 487 (1), 56-60 (2000)
MEDLINE 20584719
PUBMED 11152884
REFERENCE 3 (bases 1 to 959)
AUTHORS Genoscope.
TITLE Direct Submission
JOURNAL Submitted (07-SEP-2000) Genoscope - Centre National de Séquençage,
2 rue Gaston Crémieux, CP 5706, 91057 EVRY cedex, FRANCE. (E-mail:
sequef@genoscope.cns.fr - Web: www.genoscope.cns.fr)

```

```

COMMENT This GSS is part of a random genomic sequencing program of thirteen
yeast species: Saccharomyces bayanus var. uvarum, Saccharomyces
exiguus, Saccharomyces servazzii, Zygosaccharomyces rouxii,
Saccharomyces kluyveri, Kluyveromyces thermotolerans, Kluyveromyces
lactis var. lactis, Kluyveromyces marxianus var. marxianus, Pichia
angusta, Debaryomyces hansenii var. hansenii, Pichia sorbitophila,
Candida tropicalis and Yarrowia lipolytica. Genomic inserts of 3 to
5 kb were prepared and both extremities were sequenced. See
keywords for description of this sequence and for the sequence of
the other extremity of this insert.

FEATURES
    source
        location/Qualifiers
            1..959
            /organism="Saccharomyces kluyveri"
            /mol_type="Genomic DNA"
            /strain="CBS 3082"
            /db_xref="taxon:4934"
            /clone="AU00A005F10"
            /clone_11b="AU00A"
            /note="end: 17"
            /note="complement(<71..>958)
            /note="similar to Saccharomyces cerevisiae ORF YPR186c [
            /note="PFI1; TFI1A (transcription initiation factor) ]"
            /evidence=not_experimental

    misc_feature
        complement(<71..>958)
        /note="similar to Saccharomyces cerevisiae ORF YPR186c [
        /note="PFI1; TFI1A (transcription initiation factor) ]"
        /evidence=not_experimental

    ORIGIN
        Query Match
        Best Local Similarity 57.2%; Score 167.6; DB 9; Length 959;
        Pred. No. 6,1e-23;
        Matches 324; Conservative 1; Mismatches 235; Indels 6; Gaps 1;

Qy      211 CATTGGAAACATATTTGATATCATATTCGGAAGAAAAACCATTCATTTGCTAGTGT 270
Db      958 CACTTGGAAACATATTTGATATCATATTCGGAAGAAAAACCATTCATTTGCTAGTGT 899
Qy      271 GGTAAAGGGGTTAATTTTCGACACACTTGAAGAGACATGAATACCATTCACCAAGTCA 330
Db      898 GGTAAAGGTGTACTACAAAGAGAGATTAAGAGACATGAGATACATACCAAAATCG 839
Qy      331 TTTAAATGTACATTTGAAAATTTGTCAGAGACATTTTAAACATCAATCTTTAAAGACAT 390
Db      898 TTTAAATGTGAATATGAGGGGTGCAAGAGATTTTAAACAGATTCATCACTAAAGTCA 779
Qy      391 CATATATTTATCTGTCATGAAAACATTAACGTTAAACATGTAATTAAGTTTCACT 450
Db      778 CATACACTTTCGTCACCTTCAGAAATGACCTGTGACATTTGGGTAAATGTTCCAA 719
Qy      451 CGACCTTGAATTAAGCAACATTAATTAACATCATGAGATCTCTGCT----- 504
Db      718 AGGCTATTCGTCGCAAAACATCTTGCAGAACATCAATGTGAGATTTGAAAACAA 659
Qy      505 TATCAATGTGATCATCTCTGTTGTTTAAATTTCCAACTTGTGATTAACAATTT 564
Db      658 TATCAATGTACTTACAGGGGTGACAGAGACTTTTAAACCTGACGCTTCAAGAG 599
Qy      599 CATATATTAACATGTCATCAAAAACCTTAATGTCCTTAATGTGTAAGGTTGTTGGG 624
Db      625 AAAAAAGGTTTATCTTCATATGTTAAGTCAATGATGATTTTACATGATCAAAATATAG 684
Qy      684 GAGTCTGCTTCAGAGTACATGATGATGATGATGATGATGATGATGATGATGATGATG 479
Db      479 GAGTCTGCTTCAGAGTACATGATGATGATGATGATGATGATGATGATGATGATGATG 479
Qy      744 ACTTGATATTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 744
Db      419 AAATGTACATTTTGTGACAGAGTCTCTTTCTTAAGAGGCTGATCTTTATCTCACTAC 419
Qy      745 AATATCTTCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 770
Db      418 ATGGAACATCATTAAGATGATATATACC 393

RESULT 3
CN759437
LOCUS CN759437 801 bp mRNA linear EST 20-MAY-2004

```

DEFINITION ID0AAA25CC01RM1 ApMS Acyrthosiphon pisum cDNA clone ID0AAA25CC01 5', mRNA sequence.

ACCESSION CN759437

VERSION CN759437.1 GI:47533360

KEYWORDS EST.

SOURCE Acyrthosiphon pisum (pea aphid)

ORGANISM Acyrthosiphon pisum

REFERENCE Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Paraneoptera; Hemiptera; Sternorrhyncha; Aphidiformes; Aphidoidea; Aphididae; Macrosiphini; Acyrthosiphon.

AUTHORS 1 (bases 1 to 801)

TITLE Hunter,W., Martinez-Torres,D., Rabhe,Y., Sabater-Munoz,B., Stern,D., Tagu,D. and Wincker,P.

Unexpressed sequence tags database for the pea aphid Acyrthosiphon pisum

JOURNAL Unpublished (2004)

COMMENT Contact: D. Tagu

INRA Rennes

UMR Bio3P, BP 35327, F-35653 Le Rheu Cedex France

Tel: +33.2.23.48.51.65

Fax: +33.2.23.48.51.50

Risk of contamination by bacterial sequences from obligatory (Buchnera) or facultative endosymbionts. These sequences were obtained in the frame of the International Consortium of Aphid Genomics in collaboration with Genoscope

PCR Primers

FORWARD: CAGGAAACAGCTATGAC

Plate: 25 row: C column: 1.

FEATURES

source

1..801

Location/Qualifiers

/organism="Acyrthosiphon pisum"

/mol_type="mRNA"

/cultivar="developmentstage"

/db_xref="taxon:7029"

/clone="ID0AAA25CC01"

/tissue_type="whole insect"

/dev_stage="nymphs and adults (parthenogenetic females)"

/lab_host="XLI-Blue"

/clone_lib="ApMS"

/note="Vector: pBS-SK minus; Site 1: EcoRI; Site 2: XhoI; Sample name: ID0AAA ; Plant growth place: Department of Ecology & Evolutionary Biology, Princeton University ; Soil conditions: Soil ; Sowing date: 01/06/1999 ; Harvesting date: 01/06/1999 ; Stress date: no stress ; Description: Aphids inoculated on one-week old Vicia faba under non-sterile conditions. All parthenogenetic stages and both winged and wingless adults were collected for library construction. ; experimental condition: long photoperiod (16-hr light/8-hr dark at 18 c)"

ORIGIN

Query Match 8.4%; Score 103.8; DB 7; Length 801;

Best Local Similarity 51.5%; Pred. No. 2,7e-10;

Matches 318; Conservative 0; Mismatches 287; Indels 12; Gaps 3;

82 ACATATGAAGGGTGTATAAGCCTATATCGACCATCATTTATAGCAACATTTAGA 141

DB 8 AAATGATTTTCTGTGTAAGAAGTTACATATATACAGTTTAAACAGCATATCAAG 67

142 ACCACAGATATGATGACCGGTATTAATGTACAGTGACGATTTGTATAAGCATTTTC 201

DB 68 ACACATACCGGTGAAAAAGCGTATGATGT-----GATACGTGTGATCAAGGCTTTCT 121

202 AGAAATACATTTGGAACACATATTTGATCATTCGAAAAAAAACCATTCATTTGT 261

DB 122 TTGAATACATTTAATAGCCATTAATATGACATACCGGCGAAAGCCCTTAATATGT 181

262 TCACTGTGTGTAAGGGTTAATTCGACACACTTGAAGAAGCATGAAATCAACCAT 321

DB 182 GATTAACGTATCAAGGGTTTCTTAAGAAATCAAAATTTAAGGCAATTAAGACAT 241

322 ACAAAAGCATTTAATGTACATTTGAATTTGTCAAGACATTTTATAACATCAATCT 381

DB 242 ACCGGGAAAAAGCCGTATTAATGTGAAAAAGTAAACCAAGCGTTTCTCAGAAATCAGAT 301

QY 382 TTAAGCATCATATATATCTGTTTCAT---GAAAAACAATTACGGTAAACAATGAAT 438

DB 302 TTAATAGCTATTAATGAAGACATACCGGCGAAAGCCGTATTAATGTGATACCTGAT 361

QY 439 AAAGTTTCACTGACCTTCAAAATTAGCAACAATTAATTAAGACATCGTGTGATCT 498

DB 362 CAAGCTTTTCTCAAAAATGATGTTAATTAAGCATTAATTAAGACATACCGGTGAAGAAG 421

QY 499 CTTGCTTATCATATGATGATCATCTGCTGTTTAAAAAATTCCAAACCTGGTCATATTA 558

DB 422 C---CGTTTAAATGTGCTAATCTGTGATCAAGCGTTTATCGGAAAGCCATTTAAAAAGC 478

QY 559 CATTTTCATTAATAACACATGATCCAAAACCTTAATGTCCTTAATGTGTAAGGTGT 618

DB 479 CACTCAAGGACATATCAGGCCAAAAGCCGTATTAATGTGATACATGTAACCAAGCGTTT 538

QY 619 GTTGGGAAAAAAGTTTATCTTCACATATGTTAAGTCATGATGATTTCTACATGATCAAA 678

DB 539 TCTCAGAAATCAAAATTTTAACAGCCATACAGACACATGCCGTGAAAGCTTTTAA 598

QY 679 ATATGACTTGTGATTA 695

DB 599 TGGATTAAGCTGATTA 615

RESULT 4

CN761003 898 bp mRNA linear EST 20-MAY-2004

LOCUS ID0AAA2AE05RM1 ApMS Acyrthosiphon pisum cDNA clone ID0AAA2AE05 5', mRNA sequence.

DEFINITION CN761003

ACCESSION CN761003

VERSION CN761003.1 GI:47534926

KEYWORDS EST.

SOURCE Acyrthosiphon pisum (pea aphid)

ORGANISM Acyrthosiphon pisum

REFERENCE Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Paraneoptera; Hemiptera; Sternorrhyncha; Aphidiformes; Aphidoidea; Aphididae; Macrosiphini; Acyrthosiphon.

AUTHORS Hunter,W., Martinez-Torres,D., Rabhe,Y., Sabater-Munoz,B., Stern,D., Tagu,D. and Wincker,P.

Unexpressed sequence tags database for the pea aphid Acyrthosiphon pisum

JOURNAL Unpublished (2004)

COMMENT Contact: D. Tagu

INRA Rennes

UMR Bio3P, BP 35327, F-35653 Le Rheu Cedex France

Tel: +33.2.23.48.51.65

Fax: +33.2.23.48.51.50

Risk of contamination by bacterial sequences from obligatory (Buchnera) or facultative endosymbionts. These sequences were obtained in the frame of the International Consortium of Aphid Genomics in collaboration with Genoscope

PCR Primers

FORWARD: CAGGAAACAGCTATGAC

Plate: 2 row: E column: 5.

FEATURES

source

1..898

Location/Qualifiers

/organism="Acyrthosiphon pisum"

/mol_type="mRNA"

/cultivar="developmentstage"

/db_xref="taxon:7029"

/clone="ID0AAA2AE05"

/tissue_type="whole insect"

/dev_stage="nymphs and adults (parthenogenetic females)"

/lab_host="XLI-Blue"

/clone_lib="ApMS"

/note="Vector: pBS-SK minus; Site 1: EcoRI; Site 2: XhoI; Sample name: ID0AAA ; Plant growth place: Department of Ecology & Evolutionary Biology, Princeton University ; Soil conditions: Soil ; Sowing date: 01/06/1999 ;

Harvesting date: 01/06/1999 ; Stress date: no stress ;
Description: Aphids inoculated on one-week old *Vicia faba*
under non-sterile conditions. All parthenogenetic stages
and both winged and wingless adults were collected for
library construction ; experimental condition: long
photoperiod (16-hr light/8-hr dark at 18 °C)

ORIGIN

Query Match 8.4%; Score 103.8; DB 7; Length 898;
Best Local Similarity 51.5%; Pred. No. 2.7e-10;
Matches 318; Conservative 0; Mismatches 287; Indels 12; Gaps 3;

QY 82 ACATATGAAGGGTGTATTAAGCCCTTAATTCAGCATCATTTATGAGCAACATTAAAGA 141
DB 7 AATATGATTTCTGCTATTAAGAGTTTACTATATATCAAGTTTAAACAAGCATAAAG 66
QY 142 ACCCAAGTAATGATGACCGTATTAATGATGACGATGATGATTAAGATTTTC 201
DB 67 ACACATACCGGTGAAAAAGCCGTATGATG- - - - -GATTAAGTATCAAGCGTTTCT 120
QY 202 AGAAATACACATTGGAAACACATATTGTATCATTTCCGAAAAAAAACATTCCATTGT 261
DB 121 TGAATATCATTTATATAGCCATAAATGACACATACCGGGAAGACCTTATAATGT 180
QY 262 TCAGTGTGTGTAAGGGTTAATTCGACAACTTGAAGAGATGAATCACCAT 321
DB 181 GATTAAGTATGATCAAGCGTTTCTTAAGAAATCAAAATTTAAACAAGCATTAAGACACAT 240
QY 322 ACAAGTCAATTAATGATCATTTGAAATGTGCAAGAGCATTTATATAACATCAATCT 381
DB 241 ACCGGGAAAGCCGTATTAATGTGAAAGTGAACCAAGCGTTTCTCAAGAAATCAGAT 300
QY 382 TTAAGCATCATATATTTATCTGTTCAT- - - - -GAAAAACATTAACGTGTAACATGTAAT 438
DB 301 TTATATCGTATTAAGACACATACCGCGAAAGACCGTATTAATGATTAATCTGAT 360
QY 439 AAAGTTTCACTGACCTTCAAAATTAAGACAAACATTAATTAACATCATGATGATCT 498
DB 361 CAAGCGTTTCTCAAAATGATGTTAATTAAGCATTAATGACATATCCGTTGAAG 420
QY 499 CCGTATCATATGATGATCTCTGTTGTTTAAAAAATTTCCAACTTGTCAGATTA 558
DB 421 C- - - - -CGTTAAATGCTACTGCTGATCAAGCGTTTATCGGAAACACATTTAAAAAGC 477
QY 559 CATTTATATTAACCACTGATCCAAATTTAAATGCTTAATGTGTAATGATGATGAT 618
DB 478 CACTAGAGACACATACAGCCAAAGCCGTATTAATGATGATGATGATGATGATGAT 537
QY 619 GTTGGAAAAAGTTTATCTTCAATGATGATGATGATGATGATGATGATGATGAT 678
DB 538 TCTCAGAAATCAAAATTTAAACAAGCATTAAGAGACACATGCGGTGAAGCTGTTAA 597
QY 679 ATATGACTTGTGATTA 695
DB 598 TCGGATTAAGTGAATTA 614

RESULT 5
AK032220
LOCUS AK032220 1609 bp mRNA linear HTC 03-APR-2004
DEFINITION Mus musculus adult male olfactory brain cDNA, RIKEN full-length
homolog [Mus musculus], full insert sequence.
ACCESSION AK032220
VERSION AK032220.1 GI:26328050
KEYWORDS HTC; CAP trapper.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE
AUTHORS Carninci, P. and Hayashizaki, Y.
TITLE High-efficiency full-length cDNA cloning

JOURNAL Meth. Enzymol. 303, 19-44 (1999)
MEDLINE 99279253
PUBMED 10349636

REFERENCE

AUTHORS

Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K.,
Itoh, M., Kono, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.
Normalization and subtraction of cap-trapper-selected cDNAs to
prepare full-length cDNA libraries for rapid discovery of new genes
Genome Res. 10 (10), 1617-1630 (2000)

TITLE

Genome Res. 10 (10), 1617-1630 (2000)

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

Shibata, K., Itoh, M., Aizawa, K., Nagao, S., Sasaki, N., Carninci, P.,
Kono, H., Akiyama, J., Nishi, K., Kikunai, T., Tashiro, H., Itoh, M.,
Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishino, T., Harada, A.,
Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K.,
Fujitake, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Watanabe, M.,
Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsura, S., Kawai, J.,
Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y.
RIKEN integrated sequence analysis (RISA) system--384-format
sequencing pipeline with 384 multicapillary sequencer
Genome Res. 10 (11), 1757-1771 (2000)

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

The RIKEN Genome Exploration Research Group Phase II Team and the
PANTOM Consortium.
Functional annotation of a full-length mouse cDNA collection
Nature 409, 685-690 (2001)

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

The PANTOM Consortium and the RIKEN Genome Exploration Research
Group Phase I & II Team.
Analysis of the mouse transcriptome based on functional annotation
of 60,770 full-length cDNAs
Nature 420, 563-573 (2002)

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

Adachi, J., Aizawa, K., Akiyama, T., Arakawa, T., Bono, H., Carninci, P.,
Fukuda, S., Furumori, M., Hanagaki, T., Hara, A., Hashizume, M.,
Hayashida, K., Hayatsu, N., Hiramoto, K., Hirooka, T., Hirozane, T.,
Hori, F., Imotani, K., Ishii, Y., Itoh, M., Kagawa, I., Kasukawa, T.,
Kato, H., Kawai, J., Kojima, Y., Kondo, S., Kono, H., Kouda, M.,
Koya, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Murata, M.,
Nakamura, Y., Nishi, K., Nomura, K., Nomura, K., Ohno, M., Ohsato, N.,
Okazaki, Y., Saito, R., Saitoh, H., Sakai, C., Sakai, K., Sakazume, N.,
Sano, H., Sasaki, D., Shibata, K., Shingawa, A., Shiraki, T.,
Sogabe, Y., Tagami, M., Tagawa, A., Takahashi, F., Takaku-Akhiba, S.,
Takeda, Y., Tanaka, T., Tomaru, A., Toya, T., Yasunishi, A.,
Muramatsu, M. and Hayashizaki, Y.
Direct Submission

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

Submitted (16-JUN-2001) Yoshihide Hayashizaki, The Institute of
Physical and Chemical Research (RIKEN), Laboratory for Genome
Exploration Research Group, RIKEN Genomic Sciences Center (GSC),
RIKEN Yokohama Institute, 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama,
Kanagawa 230-0045, Japan (E-mail: genome-res@gs.c.riken.jp,
URL: http://genome.gsc.riken.jp/, Tel: 81-45-503-9222,
Fax: 81-45-503-9216)

COMMENT

CDNA library was prepared and sequenced in Mouse Genome
Encyclopedia Project of Genome Exploration Research Group in Riken
Genomic Sciences Center and Genome Science Laboratory in RIKEN.
Division of Experimental Animal Research in Riken contributed to
prepare mouse tissues.

FEATURES

source
location/Qualifiers
1..1609
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6J"
/db_xref="PANTOM:DB:6430503015"
/db_xref="taxon:10090"
/clone="6430503015"
/sex="male"

CDS

```
/cissue_type="olfactory brain"
/clone_lib="RIKEN full-length enriched mouse cDNA library"
/dev_stage="adult"
330..1223
/note="unnamed protein product; MSZF33 (FRAGMENT) homolog
[Mus musculus] (SPTR|088251, evidence: PASTY, 87.1%ID,
100%length, match=255)
putative"
/codon_start=1
/protein_id="BAC27766.1"
/db_xref="GI:26328051"
/translation="MLETYYRNLAAGVGNMEDSNIEBHCESSRRHGRHNRHTGEKPYE
GIQGEAFVHSSLSQWKKIITHEGKRYKCNQCKAVSRHSIIQIHRTSGSEKPYCN
OGKAFQSHSLKIHMTHTGKPYKCNQCKAPAGNKLQKERTHTGKPYKQKOC
NKAPEVSYQVHKHTHTGKPYKCNQCKAPARHSLKTKYHKITHGKPYKCNQKOC
ALAHSTLQVHRTHTGKPYEBCQCKAPANOSYFOVHKRHTGKPYKCDQCKAF
VSSSDLKRHRVHTGRRTLQM"
1594..1599
/note="putative"
1609
/note="putative"
```

ORIGIN

```
Query Match 7.9%; Score 98.2; DB 3; Length 1609;
Best Local Similarity 51.2%; Pred. No. 3.6e-09;
Matches 310; Conservative 0; Mismatches 283; Indels 12; Gaps 3;
```

```
QY 60 ACCTCCCAAAAGTATATTTGCACATATGAGAGGTGTGATTAAGCTTAATATGACCATC 119
DB 521 ACATACGAGAGAGAAACGTACAATATCAATATGATTAAGCTTAATATGACGACAG 580
QY 120 ATTATTAAGACCACTTTAAGAACCCGACATATGATGACGACCGTAATTAAGTACAG 179
DB 581 CATTCTACAATATATATTAAGAAACATATGATGAGAGAAACCTTAATGATGTA-----A 634
QY 180 CGATTGTATTAAGCATTTTTCAGAAATATCATTTTGAAGAACATATTTGATCATCTC 239
DB 635 TCATATGTGATTAAGCCTTTACACAAACAGTCATCTCAAAATATCATATGTTATACATAC 694
QY 240 CGAAAAAAACCATTCATTTGTCAGTGTGTGTAAGGGGTTAATTTCTGACACACCTT 299
DB 695 TGGAGAGAAACCTTACAATGTATATCAATGTGTGTAAGCCTTTCATGTCATATTAATCT 754
QY 300 GAAAAAGCATGAATATCCCATACAAAGTCATTTAAATGTATCATTTGAAAAATTTGTCA 359
DB 755 TCAAAAAACATGAAGAAACACATCTGAGAGAAACCTTAACAAGTATCATATGATATTA 814
QY 360 AGCATTT---TATTAACATCAATCTTTAAGACATCATATATATCTGTCATGAAAAAAC 416
DB 815 AGCCTTTGTATATGAAGTTATTTTACAAGTTCAATTAATAAAACACATCTGAGAGAAACC 874
QY 417 ATTAACGTGTAAACATGTATATAAGTTTTCATCTGACCTTCAAAATTAAGACACAATAA 476
DB 875 TTACAATGTATATGAATGTGTGTAAGCCTTTGACGACAGTCATCTCAAAAGTCATTA 934
QY 477 ATTAACATCATGTGTGATCTCTGCTTATCAATGTGATCATCTGTTGTTTAAAAA 536
DB 935 AATAACACATATCTGAGAGAAAC---CTTACAATGTATATCAATGTGTGTAAGCCTTTCG 991
QY 537 TTTCCAAACTTGTGCAGATTTATTAATTAATTAACAACCTGATCCAAACCTTAATG 596
DB 992 ATATCATATGATCACTCCAGATGATCAAAAGAACATCTGAGAGAAAGCCCTATGATG 1051
QY 597 TCCTAAATGTGTGAAGGTGTGTGTGGAAAAAGTTTATCTTCAATATGTTAAGTCA 656
DB 1052 TGACCAATGTGTGCAAGCCTTTGCAAAATCAAAAGTTATTTTCCAAGTACATTAAGAAATACA 1111
QY 657 TGATG 661
DB 1112 TACTG 1116
```

RESULT 6

BC047646
LOCUS BC047646 2647 bp mRNA linear HTC 19-NOV-2003
DEFINITION Homo sapiens zinc finger protein 85 (HPF4, HNF1), mRNA (cDNA clone
IMAGE:4828290), with apparent retained intron.
ACCESSION BC047646
VERSION BC047646.1 GI:29126804
KEYWORDS HTC.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens

REFERENCE

AUTHORS

Strausberg, R.L., Feingold, E.A., Grouse, L.H., Derge, J.G.,
1 (bases 1 to 2647)
Klausner, R.D., Collins, F.S., Wagner, L., Shennan, C.M., Schuler, G.D.,
Altschuler, S.F., Zeeberg, B., Buecaw, K.H., Schaefer, C.F., Bhat, N.K.,
Hopkins, R.F., Jordan, H., Moore, T., Max, S.I., Wang, J., Hsieh, F.,
Diatchenko, L., Marusina, K., Farmer, A.A., Rubin, G.M., Hong, L.,
Stapleton, M., Soares, M.B., Bonaldo, M.F., Casavant, T.L.,
Scheetz, T.E., Brownstein, M.J., Uesdin, T.B., Toshiyuki, S.,
Carninci, P., Prange, C., Raha, S.S., Loguettano, N.A., Peters, G.J.,
Abramson, R.D., Mullahy, S.J., Bosak, S.A., McEwan, P.J.,
McKernan, K.J., Malek, J.A., Gunaratne, P.H., Richards, S.,
Worley, K.C., Hale, S., Garcia, A.M., Gay, L.J., Huiy, S.W.,
Villalón, D.K., Muzny, D.M., Sodergren, E.J., Lu, X., Gibbs, R.A.,
Fahy, J., Helton, E., Kettman, M., Madan, A., Rodriguez, S.,
Sanchez, A., Whitting, M., Madan, A., Young, A.C., Shevchenko, Y.,
Bouffard, G.G., Blakeley, R.W., Touchman, J.W., Green, E.D.,
Dickson, M.C., Rodriguez, A.C., Grimwood, J., Schmutz, J., Myers, R.M.,
Butterfield, Y.S., Krzywinski, M.I., Skalska, U., Small, D.E.,
Scherer, A., Schein, J.E., Jones, S.J., and Marra, M.A.
Generation and initial analysis of more than 15,000 full-length
human and mouse cDNA sequences
Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)

TITLE

JOURNAL MEDLINE
PUBMED 12477932
2 (bases 1 to 2647)
Strausberg, R.
Direct Submission
Submitted (03-MAR-2003) National Institutes of Health, Mammalian
Gene Collection (MGC), Cancer Genomes Office, National Cancer
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
USA

REMARK
COMMENT NIH-MGC Project URL: <http://mgc.ncl.nih.gov>
Contact: MGC help desk
Email: cgabs-remail.nih.gov
Tissue Procurement: Miklos Palcovits, M.D., Ph.D.
CDNA Library Preparation: Michael J. Brownstein (NHGRI) & Shitaki
Toshiyuki and Piero Carninci (RIKEN)
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: Sequencing Group at the Stanford Human Genome
Center, Stanford University School of Medicine, Stanford, CA 94305
Web site: <http://www-shgc.stanford.edu>
Contact: (Dickson, Mark) mcd@paxil.stanford.edu
Dickson, M., Schmutz, J., Grimwood, J., Rodriguez, A., and Myers,
R. M.

FEATURES

Source

1..2647
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:4828290"
/cissue_type="Testis"
/clone_lib="NIH MGC_97"
/lab_host="DH10B"
/note="Vector: pBluescript"

ORIGIN

Query Match

7.9%; Score 97.8; DB 3; Length 2647;

Best Local Similarity 50.1%; Pred. No. 4.5e-09; Matches 330; Conservative 0; Mismatches 317; Indels 12; Gaps 3;

QY 7 GAAAGTGACGAAACCAATGCATATCATCTTTATATATCTCTCTCTTCATCAGCTCC 66
Db 1235 GAATGTGGCAAACTTTAACGATCTCTCACTCTACTACCATTAAGATATCTACT 1294
QY 67 AAAAGTATATTTGCAATATGAAAGGTGTGATTAAGCCTTAAATGACCATTAATTA 126
Db 1295 GGAGAGAAACCCATCAAAATGTAAGATGTGTAAAGCTTTTAACGATCTCAACCTT 1354
QY 127 GAGCAACATTTAAGAACCCACATTAATGATCGACCTTATTAATGTACAGGACGATGT 186
Db 1355 ACTACCATGAGAAATTCATCTGAGAGAAACCTTACAAAGT-----GAAAGATGT 1408
QY 187 GATAAAGCATTTTTCAGAAAATCACAATTGGAACAATATGTATCATCATCCGAAAA 246
Db 1409 GCGAAAGCCTTTAAGACAGTCTCAAACTTACTACATTAAGATTAATCTACTGAGAG 1468
QY 247 AAACATTCATTTGTCAGTGTGTGTAAAGGGTTAATTTCTGACACAATTGAAAA 306
Db 1469 AAACCTACAAATGTAAATAATGTGAAAAAGCTTTAACAGTCTGACACCTTACACA 1528
QY 307 CATGAATACCATTCACAAAGTCAATTAATGTACATTTGAAATTTGTCAAGAGCATTT 366
Db 1529 CATGAGATTAATCTACTGAGAGAAACCTTACAAATGTGAAAAATGTGAAAAAGCTTT 1588
QY 367 TATPA---ACATCAATCTTTAAGACATCATATATATCTGTCATGAAAAACATTAAG 423
Db 1589 AATCATTTCTCACACCTTACTACATTAAGATTAATCTACTGAGAGAAACCTTACAA 1648
QY 424 TGTAAACATGTAAAGTTTCACTCGACCTTCAAAATTAACACACATTAATTA 483
Db 1649 TGTAAAGATGTGTAAAGCTTTTAAACACTTTCAACCTTACTAAACATTAAGATTA 1708
QY 484 CATCATGTGTGATCTCTGCTTATCATGTGATCATCTGCTGTTTAAAAATTTCGA 543
Db 1709 CATACGTGAGAGAGAGCTTACAAATGTAAAGATGTGAAAAAGCTTTTAAACCAATCTCA 1768
QY 544 ACTGTGCTAGTATTAACATTTTCAATTAACAACTGATCCAAACCTTAATGCTTAA 603
Db 1769 AAACCTACTGACATTAAGAAATTCATCTGAGAGAAACCTTA---TGAATGTGAAAA 1825
QY 604 TGTGTAAAGGTGTGTGGGAAAAAGTTTATCTTCAATATGTTAAATCATGATGA 662
Db 1826 TGTGCAAAAGCTTTTAAACGAGCTCTCAAAATCTTACTAGACATTAAGAAAAAGTCA 1884

RESULT 7
AUI23448 745 bp mRNA linear EST 01-AUG-2002
LOCUS AUI23448 NT2RM2 Homo sapiens cDNA clone NT2RM2000319 5', mRNA
DEFINITION
sequence.
ACCESSION AUI23448
VERSION AUI23448.1 GI:10948164
KEYWORDS
SOURCE EST.
ORGANISM Homo sapiens (human)
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
REFERENCE
AUTHORS 1 (bases 1 to 745)
Ota, T., Wakamatsu, A., Ozawa, M., Ishii, S., Saito, K., Yamamoto, J.,
Nakamura, Y., Nishikawa, T., Nagai, T., Suzuki, Y., Sugano, S. and
Isogai, T.
TITLE HRI human cDNA project (Ota, T., Wakamatsu, A., Ozawa, M., Ishii, S.,
Saito, K., Yamamoto, J., Nakamura, Y., Nishikawa, T., Nagai, T.,
Suzuki, Y., Sugano, S., Isogai, T.)
JOURNAL Unpublished (2000)
COMMENT Contact: Takao Isogai
Genomics Laboratory
Helix Research Institute
1532-3 Yama, Kisarazu, Chiba 292-0812, Japan
Tel: 81-438-52-3975

Fax: 81-438-52-3986
Email: genomics@hri.co.jp
HRI human cDNA project; 5'- & 3'-end one pass sequencing: Helix
Research Institute; cDNA library construction: Department of
Virology, Institute of Medical Science, University of Tokyo, and
Helix Research Institute.

FEATURES

source
1..745
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="NT2RM2000319"
/cell_type="teratocarcinoma"
/cell_line="NT2"
/clone_id="NT2RM2"
/note="Vector: pME18SFL3; mRNA from uninduced NT2 neuronal
precursor cells"

ORIGIN

Query Match 7.9%; Score 97.6; DB 1; Length 745;
Best Local Similarity 50.2%; Pred. No. 4.5e-09;
Matches 328; Conservative 0; Mismatches 314; Indels 12; Gaps 3;

QY 7 GAAAGTGACGAAACCAATGCATATCATCTTTATATATCTTCTTCTTCATCAGCTCC 66
Db 87 GAATGTGGCAAACTTTTAAACGATCTCACTTACTACCATTAAGATTAATCTACT 146
QY 67 AAAAGTATATTTGCAATATGAAAGGTGTGATTAAGCCTTAAATGACCATTAATTA 126
Db 147 GAGAGAAACCTTACAAATGTAAAGATGTGTAAAGCTTTTAAACGCTTCTTCAACCTT 206
QY 127 GAGCAACATTTAAGAACCCACATTAATGATCGACCTTAAATGTACAGTGAAGCATTT 186
Db 207 ACTACCATTAAGAAATTCATCTGAGAGAAACCTTCAAAATGT-----GAAAGATGT 260
QY 187 GATAAAGCATTTTTCAGAAAATCACAATTGGAACAATATGTATCATCATCCGAAAA 246
Db 261 GCGAAAGCCTTTAAGACAGTCTCAAACTTACTACATTAAGATTAATCTACTGAGAG 320
QY 247 AAACATTCATTTGTCAGTGTGTGTAAAGGGTTAATTTCTGACACAATTGAAAA 306
Db 321 AAACCTACAAATGTAAATAATGTGAAAAAGCTTTTAAACAGTCTGACACCTTACTACA 380
QY 307 CATGAATACCATTCACAAAGTCAATTAATGTACATTTGAAATTTGTCAAGAGCATTT 366
Db 381 CATGAGTAAATTCATACGTGAGAGAAACCTTACAAATGTGAAAAATGTGAAAAAGCTTT 440
QY 367 TATPA---ACATCAATCTTTAAGACATCATATATATCTGTTCATGAAAAACATTAAG 423
Db 441 AATCATTTCTCACACCTTACTACATTAAGATTAATCTACTGAGAGAAACCTTACAA 500
QY 424 TGTAAACATGTAAAGTTTCACTCGACCTTCAAAATTAAGCAACATTAATTA 483
Db 501 TGTAAAGATGTGTGTAAGCTTTTAAACACTTCTTAAACCTTACTTAAACATTAAGATTA 560
QY 484 CATCATGTGTGATCTCTGCTTATCAATGATGATCATCTGCTGTTTAAAAATTTCGA 543
Db 561 CATACTGAGAGAAACCTTACAAATGTAAAGATGTGAAAAAGCTTTTAAACCAATCTCA 620
QY 544 ACTGTGCTAGTATTAACAATTCATATTAACAACATGCAATCCAAAACCTTAATGCTTAA 603
Db 621 AAACCTACTGACATTAAGAAATTCATCTGAGAGAAACCTTA---TGAATGTGAAAA 677
QY 604 TGTGTAAAGGTGTGTGGGAAAAAGTTTATCTTCAATGTTAAATGATGAT 657
Db 678 TGTGCAAAAGCTTTTAAACGAGCTCTCAAAATCTTACTAGACATTAAGAAAAAGTCA 731

RESULT 8
CO434402 727 bp mRNA linear EST 06-JUL-2004
LOCUS CO434402
DEFINITION UI-M-HX0-csl-h-09-0-UI.r1 NIH_BMAP_HX0 Mus musculus cDNA clone
IMAGE:30687440 5', mRNA sequence.

ACCESSION CO434402
VERSION CO434402.1 GI:49680696
KEYWORDS EST.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 727)
NIH-MGC <http://mgs.nci.nih.gov/>
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
JOURNAL
COMMENT
Contact: Robert Strausberg, Ph.D.
Email: cgabbs@mail.nih.gov
Tissue Procurement: Dr. James Lin University of Iowa
CDNA Library Preparation: Dr. M. Bento Soares, University of Iowa
CDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
Clone Distribution: Distribution information can be found at
<http://genome.uiowa.edu/distribution/mouseefl.html>
This clone was contributed by the Brain Molecular Anatomy Project
(BMAP)
Seq primer: pyx-5.
FEATURES
source
Location/Qualifiers
1..727
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6"
/db_xref="taxon:10090"
/clone="IMAGE:30687440"
/issue_type="whole eye"
/dev_stage="newborn (1, 5, 15 days) and embryonic (15, 16,
17, 18 dpc)"
/lab_host="DH10B (T1 phage resistant)"
/clone_lib="NIM_BMAP_HX0"
/note="Organ: Eye; Vector: pyx-Asc; Site 1: Ecor I;
Site 2: Not I; The library was constructed according
Bonaldo, Lennon and Soares, Genome Research, 6:791-806,
1996. Denatured RNA was size fractionated on a 1% agarose
gel. First strand cDNA synthesis was primed with oligo-dT
primer containing a Not I site. Double strand cDNA was
size selected according to mRNA size fraction, ligated
with Ecor I adaptor, digested with NotI and then cloned
directionally into pyx-Asc vector. The library tag
sequence located between the Not I site and the polyA tail
is AATTAATTACG. This library was created for the University
Iowa Brain Anatomy Project (BMAP): 'Gene Discovery in the
Developing Mouse Nervous System', supported by National
Institute of Mental Health (NIMH)."

ORIGIN
Query Match 7.5%; Score 93.4; DB 7; Length 727;
Best Local Similarity 52.1%; Pred. No. 3e-08;
Matches 261; Conservative 0; Mismatches 231; Indels 9; Gaps 2;

4 AGTGAAGAGGAGCAACCAATGATATCATCTTTAATCTCTCTTCATACCGT 63
DB AATCAGTGTGATTAAGCTTTTCTCGATACAGTACTTCAATCTAATGAGAACA 137
64 CCCAAAAGATATTTTGACATATGAAGGTGTGATAAGCCTATATGACCATATTA 123
DB ACTGAGAGAGAAACCTTCAATGTATCACTGTGATTAAGCCTTTCTGAGAAATGAT 197
124 TTGAGCAACATTTAAGAACCCACAGTATGATGACCGTATTAATGTATGACGAT 183
DB CTCCAAAACCATGTAAAGATACATCTGAGAGAAACCCCTACAAATGTA-----ATCAA 251
184 TGTGATTAAGCATTTTTCAGAAATACATTTTGAACAACATTTGATTCATTCGAA 243
DB TGTGATTAAGCCTTTTCCCTTAATCTACCTCCACATTTATGAAAGACATACCGG 311
244 AAAAAACATTCATTTTCAGTGTGTGTGTAAGGGGTTAATTCGACAAACATTTGAA 303
DB GAAAAACCCATCAAAATGATATCACTGTGATTAAGCCTTTCCGATACATACCTCC 371

304 AGACATGAATGATACCCATACAAAGTCATTTAATGTACATTTGAATTTGCAAGA 363
DB ATTCATAGAGAAACACATATCTGAGAGAAACCTTACAAATGTATATCATGTAT 431
364 TTTTATTAACATCAATCTTTAAGACATCATATATATC---TGTTCATGAAAAACATTA 420
DB TTTTTCATATACCTTTTCTCTTCAAACTCACAGAAAGACATATCTGAGAAACCATTC 491
421 AGCTGAACATGTATTAAGTTTCACTCGACCTTCAAAATTTAGCACATTAATTA 480
DB AATGTAAATCAATGTATTAAGTCTTTCTCAATACAGTATCTTCAAACTCATAGA 551
481 AAACATCATGTGATCTCCT 501
DB 552 ACACATCTGAGAGAAACCT 572

RESULT 9
LOCUS CN352904 631 bp mRNA linear EST 16-MAY-2004
DEFINITION 17000600057466 GRN_PREHEP Homo sapiens cDNA 5', mRNA sequence.
ACCESSION CN352904
VERSION CN352904.1 GI:47352838
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 631)
REFERENCE
AUTHORS Brandenberger, R., Wei, H., Zhang, S., Lei, S., Murage, J., Fisk, G.J.,
Li, Y., Xu, C., Pang, R., Guegler, K., Rao, M.S., Mandalam, R.,
Lebkowksi, J. and Stanton, L.W.
Transcriptome characterization elucidates signaling networks that
control human ES cell growth and differentiation
Nat. Biotechnol. 22 (6), 707-716 (2004)
CONTACT: Brandenberger R
Regenerative Medicine
Genon Corporation
230 Constitution Drive, Menlo Park, CA 94025, USA
Tel: 650 473 8658
Fax: 650 473 7760
Email: rbrandenberger@genon.com
Insert Length: 631 Std Error: 0.00.
FEATURES
source
Location/Qualifiers
1..631
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/issue_type="embryonic stem cells, DMSO-treated H9 cell
line"
/clone_lib="GRN_PREHEP"
/note="oligo dT primed, full-length enriched cDNA library
from DMSO-treated hES cell line H9 (p22) maintained in
feeder-free conditions"

ORIGIN
Query Match 7.5%; Score 92.4; DB 7; Length 631;
Best Local Similarity 54.1%; Pred. No. 4.7e-08;
Matches 236; Conservative 0; Mismatches 191; Indels 9; Gaps 2;

69 AAAGTATATTTGACATATGAAGGGTGTGATAAGCCTTAATCGACCATATTAATA 128
DB AAGGAAGCCTTCAAAATGTAAAGTAAGAAAAAGCTTTTAACCAATCTCAAAACTTAC 142
129 GCAACATTTAAGAACCCACAGTAAATGATGACCGTATTAATGTATACAGTGA 188
DB TGAACATTAAGAAATTCATCTGAGAGAAACCCATATGATGT-----GAAAAATGTG 196
189 TTAAGCATTTTTCAGAAAAATCAATTTGAAAAACATATTTATATCATTCGAAAA 248
DB CAAAGCTTTTAACAAATCTCTCAAAATCTTACTGACATTAAGAAAAAGTCATACAG 256

QY 249 ACCATTCCATTGTCAGTGTGTGTAAGGGGTTAATTCGACAACTTGAAAAGACA 308
Db 257 ACCCTTCAAAATGTGAAGAAATGGCAAGTTTAAATGGCCCTCAACCTTACTATCA 316
QY 309 TGAATACCCATGACAAAGTCATTTAAATGTCATTTGAATGTGCAGAAAGACATTTA 368
Db 317 TAAAGATTAATTCATACGTGAGAGAAACCATTAACAATGTGAAGATGTGGCAAAAGCTTTAA 376
QY 369 TAAA---CATCATCTTTAAGACATCATATATATCTGTTCAATGAAAAAATTAACGTG 425
Db 377 CCAATCCTCAAAACTTAACCAACATTAAGAAATTCATCTGAGAGAAACCTCAGCAG 436
QY 426 TAAACATGATTAATAAGTTTCACTGACCTTCAAAATTAGCAACAATTAATTAACA 485
Db 437 TGAAGATGTGGCAAAAGCCTTTAACCAGTCTCAAACTTACTAATTAAGATGATTCGA 496
QY 486 TCATGTGTGATCTCCT 501
Db 497 TACTGGAGAAAAAACCCT 512

RESULT 10
LOCUS BC036394 2672 bp mRNA linear HTC 19-NOV-2003
DEFINITION Homo sapiens zinc finger protein 85 (HPF4, HTP1), mRNA (cDNA clone IMAGE:5259399), containing frame-shift errors.
ACCESSION BC036394
VERSION BC036394.1 GI:23025784
KEYWORDS HTC.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE 1 (bases 1 to 2672)
Strausberg, R.L., Ringold, P.A., Grouse, L.H., Derge, J.G., Klausner, R.D., Collins, F.S., Wagner, L., Shenmen, C.M., Schuler, G.D., Altschuler, S.F., Zeeberg, B., Buetow, K.H., Schaefer, C.F., Bhat, N.K., Hopkins, R.F., Jordan, H., Moore, T., Max, S.I., Wang, J., Hsieh, F., Hopenhaver, L., Marusina, K., Farmer, A.A., Rubin, G.M., Hong, L., Stapleton, M., Soares, M.B., Bonaldo, M.F., Casavant, T.L., Scheetz, T.E., Brownstein, M.J., Ueda, T.B., Toshitoki, S., Carninci, P., Prange, C., Raha, S.S., Loguettano, N.A., Peters, G.J., Abramson, R.D., Mulhally, S.J., Bosak, S.A., McEwan, P.J., McKernan, K.J., Malek, J.A., Gunaratne, P.H., Richards, S., Worley, K.C., Hale, S., Garcia, A.M., Gay, L.J., Hult, S.W., Villalón, D.K., Muzny, D.M., Sodergren, E.J., Lu, X., Gibbs, R.A., Sanchez, J., Helton, E., Kettelman, M., Madan, A., Rodriguez, S., Bouffard, G.G., Blakeley, R.W., Touchman, J.W., Green, E.D., Dickson, M.C., Rodriguez, A.C., Grimwood, J., Schmutz, J., Myers, R.M., Butlerfield, Y.S., Krzywinski, M.I., Skalska, U., Smal, D.B., Scherch, A., Schein, J.E., Jones, S.J. and Marra, M.A.
Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences
Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)

TITLE 2 (bases 1 to 2672)
JOURNAL Strausberg, R.
MEDLINE Direct Submission
PUBMED Submitted (05-AUG-2002) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA
REFERENCE NIH-MGC Project URL: <http://mgc.nci.nih.gov>
COMMENT Contact: MGC help desk
Email: cgabbs-r@mail.nih.gov
Tissue Procurement: Miklos Palkovits, M.D., Ph.D.
CDNA Library Preparation: Michael J. Brownstein (NHGRI) & Shizuki Tohshiki and Piero Carninci (RIKEN)
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: Institute for Systems Biology
<http://www.systemsbiology.org>

contact: amadansystembiology.org
Anup Madan, Jessica Fahy, Erin Helton, Mark Kettelman, Anuradha Madan, Stephanie Rodrigues, Amy Sanchez and Michelle Whiting
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at: <http://image.llnl.gov>
Series: IPAK Place: 73 Row: a Column: 16
This clone has the following problem: frame shifted.
FEATURES
source
1..2672
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:5259399"
/issue_type="Brain, hippocampus"
/clone_id="NIH_MGC_95"
/lab_host="DH10B"
/note="Vector: pBluescript"

Query Match 7.5%; Score 92.4; DB 3; Length 2672;
Best Local Similarity 54.1%; Pred. No. 5.3e-08;
Matches 236; Conservative 0; Mismatches 191; Indels 9; Gaps 2;
QY 69 AAGATTAATTTGACATATGAAGGGTGTGATTAAGCTTAATATGACCATCATTAATAGA 128
Db 1745 AGAGAAAGCCTTAACAATGTAAAGATGTGAAGAAAGCTTTTAACCAATCCTCAAACTTAC 1804
QY 129 GCAACATTTAAGAACCCACAGTAATGATGACACCGTATTAATGTACAGTGACGATTTGA 188
Db 1805 TGAACATTAAGAAATTCATCTACGTGAGAGAAACCTCATTAATGT-----GAAAAATGTGG 1858
QY 189 TAAACATTTTTCAGAAATATCATTTTGAACATATTTGATATCATATTCGAAAAA 248
Db 1859 CAACGCTTTTACCAAGCTCTCAATCTTACATAGATTAAGAAATCATAGAGAGAA 1918
QY 249 ACCATTCATTTGTCAGTGTGTGTAAGGGGTTAATTCGACAACTTGAAAAACA 308
Db 1919 ACCTTACAAAGTGAAGATGTGGCAAGGTTTAAATGGCCCTCAACCTTACTATCA 1978
QY 309 TGAATTCACCATCAAAAGTCATTTAATGTATGATTTGAAAAATTCAGAAAGCATTTTA 368
Db 1979 TAAATTAATTCATCTACGTGAGAGAAACCATTAAGTAAGATGTGGCAAGCTTTTAA 2038
QY 369 TAAA---CATCATCTTTAAGACATCATATATATCTGTTCAATGAAAAAATTAACGTG 425
Db 2039 CCAATCCTCAAAACTTACCAAACTTAAGAAATTCATCTGAGAGAAACCTCAGCAG 2098
QY 426 TAAACATGATTAATAAGTTTCACTGACCTTCAAAATTAGCAACAATTAATTAACA 485
Db 2099 TGAAGATGTGGCAAAAGCCTTTAACCAGTCTCAAACTTACTAATTAAGATGATTCGA 2158
QY 486 TCATGTGTGATCTCCT 501
Db 2159 TACTGGAGAAAAAACCCT 2174

RESULT 11
LOCUS A0005136 591 bp DNA linear GSS 27-JUN-1998
DEFINITION CIT-HSP-2290017.TF CIT-HSP Homo sapiens genomic clone 2290017, genomic survey sequence.
ACCESSION A0005136
VERSION A0005136.1 GI:3082581
KEYWORDS GSS.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE 1 (bases 1 to 591)
Adams, M.D., Rounley, S.D., Zhao, S., Field, C.E., Bass, S., Linher, K., Golden, K., Berry, K., Granger, D., Sun, E., Wible, C., Shizuya, H., Simon, M. and Venter, J.C.

TITLE Use of a random BAC End Sequence Database for Sequence-Ready Map Building (1998)
JOURNAL Unpublished (1998)
COMMENT Other_GSSs: CIT-HSP-2290017.TR
Contact: Mark Adams
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: mdamad@igf.org

Clones are available from Research Genetics (info@resgen.com). BAC end search page:
http://www.tigr.org/tdb/humgen/bac_end_search/bac_end_search.html
Seq primer: M13-21;
Class: BAC ends.

FEATURES

source
1..591
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="GDB:7150844"
/db_xref="taxon:9606"
/clone="2290017"
/sex="Male"
/cell_type="Sperm"
/clone_lib="CIT-HSP"
/note="Vector: pBelobAC11; Site_1: HindIII; Site_2: HindIII"

ORIGIN

Query Match 7.4%; Score 92; DB 8; Length 591;
Best Local Similarity 52.2%; Pred. No. 5,7e-08;
Matches 256; Conservative 0; Mismatches 225; Indels 9; Gaps 2;
Db 7 GAAAGTACGAAACCAATGATATCATCTTTAATATCTTCTCTTCATCAGCGCC 66
69 GAATGTGGCAAGCCTTCTACCACTTCTTACACCTTACACATTAAGTAATTCATCT 128
Qy 67 AAAAAGTATTTTGACATATGAAAGGTGTGATTAAGCCATTAATGACCATCATTA 126
129 GGAAGAAAGCCCTTCAATGTGAAGATGTGTAAGCCTTTTAACCAACCTTCAGCCCT 188
Db 127 GAGCAATTTAAGAACCCAGTAATGATCGAACCGTAAATGTAAGTGAAGCATTTG 186
189 ACTACACATTAAGTTCATCTTCACTTTAAGAAAAACCTTCAAAATGT-----GAAGATGT 242
Qy 187 GATTAAGCATTTTTCAGAAAAATCACATTTGAAAAACATATTGTATCACATTCGAAA 246
243 GACAAAGCTTTTAAACGATTTCTATACCTTAACATTAAGTAATTAATTCATTCGAGAG 302
Db 247 AAACCATTCATTTGTCAGTGTGTGTAAGGGGTTAATTCGACAAACATTGAAAAGA 306
303 AAATCTTCAAAAGTGAACATGTGGCAAGGCTTTAACTGCTTCACACCTTACAAA 362
Qy 307 CATGAATCACCCATCAAAAGTCATTTAAATGTACATTTGAAAATTGTCAAGAGCATTT 366
363 CATTAAGAAATTAATCTGAGAGAAACCTTCAAAATGTAAAGAAATGTGCAAAAGCTTT 422
Db 367 TATTAACAT--CAATCTTTAAGACATCATATTAATCTGTCATGAAAAAACATTACG 423
423 AATGTGCTTACACCTTACTACATTAAGATGATTCATCTGGAAGAAACCTTACAA 482
Qy 424 TGTAAACATGTATTAAGTTTCACTGACCTTCAAAATTAAGCAACATTAATTAAA 483
483 TGTGAAGAAATGTGCAAGCCTTTAAACACTCTCTCAAACTTAATAATTAAGATTAAT 542
Db 484 CATCATGTGTG 493
543 CATCTGGAG 552

RESULT 12
CN362485

LOCUS CN362485 687 bp mRNA linear EST 16-MAY-2004
DEFINITION 17000424188118 GRN_EB Homo sapiens cDNA 5', mRNA sequence.
ACCESSION CN362485
VERSION CN362485.1 GI:47362419
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens

REFERENCE 1 (bases 1 to 687)
Brandenberger, R., Wei, H., Zhang, S., Lei, S., Murege, J., Fiek, G., J.,
Li, Y., Xu, C., Pang, R., Guegler, K., Rao, M. S., Mandalam, R.,
Lebkowski, J. and Stanton, L. W.
Transcriptome characterization elucidates signaling networks that
control human ES cell growth and differentiation
Nat. Biotechnol. 22 (6), 707-716 (2004)

CONTACT: Brandenberger R
Regenerative Medicine
Genon Corporation
230 Constitution Drive, Menlo Park, CA 94025, USA
Tel: 650 473 9658
Fax: 650 473 7760
Email: rbrandenberger@genon.com

FEATURES

source
1..687
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/cissue_type="embryonic stem cells, embryoid bodies
derived from H1, H7 and H9 cells"
/clone_lib="GRN_EB"
/note="ToIigo dT primed, full-length enriched cDNA library
from embryoid body outgrowth derived from hES cell lines
H1 (p32), H7 (p29), and H9 (p26) maintained in feeder-free
conditions."

ORIGIN

Query Match 7.4%; Score 92; DB 7; Length 687;
Best Local Similarity 52.2%; Pred. No. 5,7e-08;
Matches 256; Conservative 0; Mismatches 225; Indels 9; Gaps 2;
Db 7 GAAAGTACGAAACCAATGATATCATCTTTAATATCTTCTCTTCATCAGCGCC 66
183 GAATGTGGCAAGCCTTCTACCACTTCTTACACCTTACACATTAAGTAATTCATCT 242
Qy 67 AAAAAGTATTTTGACATATGAAAGGTGTGATTAAGCCATTAATGACCATCATTA 126
243 AGATGAATTTTCAAAATGTGAAGCATTAAGAAAGCCTTTAACTGCTTCACACCTT 302
Db 127 GAGCAATTTAAGAACCCAGTAATGATCGACGTAATTAATGTACAGTGAAGCATTTG 186
303 AAATCAATTAAGAAATTAATCTGAGAAAAACCTTACAAATGTAAA-----GATGT 356
Qy 187 GATTAAGCATTTTTCAGAAAAATCACATTTGAAAAACATATTGTATCACATTCGAAA 246
357 GCGAAAGCCTTTTACCAAGCCTTACACCTTATTAAGCATTAAGAAATTCATCTGAAG 416
Db 247 AAACCATTCATTTGTCAGTGTGTGTAAGGGGTTAATTCGACAAACATTGAAAAGA 306
417 AAACCTTAAGAAATGTGAAGATGTGCAAAAGCCTTTAAACAGATCAAGACCTTACACA 476
Qy 307 CATGAATCAACCATCAAAAGTCATTTAAATGTACATTTGAAAATTGTCAAGAGCATTT 366
477 CATTAATTAATTAATCTGAGGAAATTCCTCAAAATGTGAAGAAATGTGTAGAGCTTT 536
Db 367 TATTAACATCAATCTTTAAGACATCATTAATTAATCTGTT-----CATGAAAAACATTACG 423
537 AACCAAGCTCAAGCTTACTGAAACATTAATTAATTAATCAACGAGAGAAACCTTATGAA 586
Qy 424 TGTAAACATGTATTAAGTTTCACTGACCTTCAAAATTAAGCAACATTAATTAAA 483
597 TGTGAAGAAATGTGCAAGCCTTTTAAACGATCTCTCAAACTTAATAATTAAGATTAAT 656

QY 484 CATCATGTG 493
DB 657 CATACTGGAG 666

RESULT 13
CN417582 752 bp mRNA linear EST 16-MAY-2004
LOCUS 17000424010604 GRN_EB Homo sapiens cDNA 5', mRNA sequence.
DEFINITION CN417582
ACCESSION CN417582
VERSION CN417582.1 GI:47405176
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
1 (bases 1 to 752)
Brandenberger, R., Wei, H., Zhang, S., Lei, S., Muraige, J., Fisk, G.J.,
Li, Y., Xu, C., Fang, R., Guegler, K., Rao, M.S., Mandalam, R.,
Lebowicki, J. and Stanton, L.W.
Transcriptome characterization elucidates signaling networks that
control human ES cell growth and differentiation
Nat. Biotechnol. 22 (6), 707-716 (2004)
Contact: Brandenberger R
Regenerative Medicine
Gen Corporation
230 Constitution Drive, Menlo Park, CA 94025, USA
Tel: 650 473 8658
Fax: 650 473 7760
Email: rbrandenberger@genon.com
Insert Length: 752 Std Error: 0.00.
Location/Qualifiers
1..752
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/issue_type="embryonic stem cells, embryoid bodies
derived from H1, H7 and H9 cells"
/clone_lib="GRN EB"
/note="oligo dt primed, full-length enriched cDNA library
from embryoid body outgrowth derived from h9 cell lines
H1 (p32), H7 (p29), and H9 (p26) maintained in feeder-free
conditions."

ORIGIN
Query Match 7.4%; Score 92; DB 7; Length 752;
Best Local Similarity 52.0%; Pred. No. 5.8e-08;
Matches 259; Conservative 0; Mismatches 230; Indels 9; Gaps 2;

QY 7 GAAAGTGAAGAAACCAATGATGATCATCTTTAATATCTTCTTCATCAGCGCC 66
DB 72 GAATGGGAGAAAGCCTTACCATCTTTCACACCTTACATACATAAGTAATTCATCT 131
QY 67 AAAAGTATATTGTCATATGAAAGGCTGTATAAAGCCTATTAAGACCATCATTTATTA 126
DB 132 GGAGAGAAAGCCTTCAATAGTGAAGATGTGTAAGCTTTTAAACACCTTCAGCCCTT 191
QY 127 GAGCAACATTTAAGAACCCAGATATGATGACCCGTATTAATGTACATGAGCATTTCT 186
DB 192 GCTACACATAGTTCATTCATGTTAAAGAAAAACCTTACAAATGT-----GAGGATGT 245
QY 187 GATTAAGCATTTTTCAGAAATCACAATTTGAAAAACATATTGTATCATATTCGGAATA 246
DB 246 GACAAAGCTTTTAAACGATTTCTATCTTACATTAACATTAAGATTAATTCATTTGAGAG 305
QY 247 AAACATTCATTTTCACTGTGTGTAAAGGGGTTAATTCGACACACATTTGAAAGA 306
DB 306 AATCTTACAAATGTGAACAATGTGGCAAGGCTTAACTGCTTCAACCCCTTACAAA 365
QY 307 CATGAATACCCATACAAAGTCAATTTAATGTACATTTGAAATTTGTCAAGAAGATTT 366
DB 366 CATAGAAATTCATCTAGAGAAAGAAACCTTACAAATGTGAAGATGTGCAAAAGCTTT 425

QY 367 TATAAACAT---CAATCTTAAAGACATATATATATCTGTTTCATGAAAAACATTAAG 423
DB 426 AATGTGCTTCACACCTTACTACATAGATGATTCATCTAGAGAAACCTTACAAA 485
QY 424 TGTAAACATGTATTAAGAGTTTTCATGCTGACCTTCAAAATTTAGCAACATTAATTTAAA 483
DB 486 TGTGAAGATGTGGCAAGGCCCTTTAACCATCTCTCAAAACCTTATCATTAAGATTAAT 545
QY 484 CATCATGGTGATCTCTCT 501
DB 546 CATACTGGAGAGAAACCT 563

RESULT 14
B99387 593 bp DNA linear GSS 26-JUN-1998
LOCUS CIT-HSP-2281A10.TF CIT-HSP Homo sapiens genomic clone 2281A10,
DEFINITION genomic survey sequence.
ACCESSION B99387
VERSION B99387.1 GI:3027197
KEYWORDS GSS.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
1 (bases 1 to 593)
Adams, M.D., Rounnley, S.D., Zhao, S., Field, C.E., Bass, S., Linher, K.,
Simon, M., and Venter, J.C.
Use of a random BAC End Sequence Database for Sequence-Ready Map
Building (1998)
Unpublished (1998)
Contact: Mark Adams
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: mdamas@tigr.org
Clones are available from Research Genetics (info@reagen.com). BAC
end search page:
http://www.tigr.org/cdb/hungen/bac_end_search/bac_end_search.html
Seq primer: M13-21;
Classes: BAC ends.

FEATURES
source location/Qualifiers
1..593
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="GDB:7147044"
/db_xref="taxon:9606"
/clone="2281A10"
/sex="Male"
/cell_type="Sperm"
/clone_lib="CIT-HSP"
/note="Vector: pBelobAC11; Site_1: HindIII; Site_2:
HindIII"

ORIGIN
Query Match 7.3%; Score 91; DB 8; Length 593;
Best Local Similarity 54.4%; Pred. No. 8.9e-08;
Matches 231; Conservative 0; Mismatches 185; Indels 9; Gaps 2;

QY 69 AAAGTATTTTGCATATGAAAGGCTGTGATTAAGCCTATATGACCATCATTTATTA 128
DB 47 AAAGAAACCTTACAAATGTGAAGAAATGTGGCAAGCTTTTAAAGCAGTCTCAACCTTAG 106
QY 129 GCAACATTTAAGAACCCAGATATGATGACCCGTATTAATGTACAGTGAAGATTTGTA 188
DB 107 AAAACATGATTAATTCATCTGAGAGAAACCTTACAAATGT-----GAGGATGTGG 160
QY 189 TAAACATTTTTCAGAAATTCACATTTGAAACACATATTTGTATCATATTCGAAAAAA 248

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 8, 2005, 15:03:15 / Search time 73 Seconds
(without alignments)
2182.812 Million cell updates/sec

Title: US-09-831-804-3

Perfect score: 2229
Sequence: 1 MESSDETKSISSLSISSSSSS.....PLVKARMDLPNETSVISR 412

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 2105692 seqs, 386760381 residues

Total number of hits satisfying chosen parameters: 2105692

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database : A_Geneseq_16Dec04:*
1: geneseqp1980s:*
2: geneseqp1990s:*
3: geneseqp2000s:*
4: geneseqp2001s:*
5: geneseqp2002s:*
6: geneseqp2003as:*
7: geneseqp2003bs:*
8: geneseqp2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2229	100.0	412	3	AA93316 A transcr
2	2225	99.8	412	5	ABP73656 Candida a
3	445	20.0	564	6	ABJ2655 Aspergill
4	445	20.0	564	6	ABJ26255 Aspergill
5	416.5	18.7	1212	4	ABG00399 Novel hum
6	407	18.3	664	8	ADN9509 Novel hum
7	403	18.1	1230	4	AAU30831 Novel hum
8	400	17.9	404	7	ADMO5916 Human pro
9	400	17.9	803	8	ADN06034 Antipepti
10	400	17.9	809	4	AAW38689 Human pol
11	399	17.9	444	8	ADN9691 Novel hum
12	398	17.9	444	8	ADN9690 Novel hum
13	398	17.9	487	8	ABO60215 Human gen
14	398	17.9	563	6	ABU96690 Human nuc
15	398	17.9	563	6	ADU60246 Secreted
16	397.5	17.8	423	8	ADQ29657 Human col
17	397.5	17.8	423	8	ADP24748 PRO polyp
18	395.5	17.7	363	3	AA93317 A human t
19	395.5	17.7	409	4	AA975181 Human col
20	395.5	17.7	423	2	AA91305 Transcrip
21	395.5	17.7	1191	7	ADJ70299 Human hea
22	395	17.7	622	4	AAW78947 Human pro
23	395	17.7	631	4	AAW79931 Human pro
24	395	17.7	632	4	ABG18386 Novel hum
25	395	17.7	719	4	ABG16953 Novel hum

26	393.5	17.7	1191	8	ADN04208 Antipepti
27	393.5	17.7	1472	4	AAU31578 Novel hum
28	393	17.6	517	4	ABG03375 Novel hum
29	392.5	17.6	1050	4	ABG09685 Novel hum
30	392.5	17.6	1050	7	ADG32995 Human nov
31	392	17.6	555	6	ADA54763 Human pro
32	391	17.5	892	7	ADMO4984 Human pro
33	390.5	17.5	878	4	AAU29528 Human pro
34	390.5	17.5	1520	4	ABG27130 Novel hum
35	390	17.5	540	7	ADG55594 Human pro
36	390	17.5	576	6	ADA54797 Human pro
37	390	17.5	620	6	ABU96725 Human nuc
38	389	17.5	891	8	ADQ67674 Novel hum
39	388.5	17.4	501	4	ABH71369 Drosophi1
40	388	17.4	568	5	ABH79872 TRAF6-inh
41	387.5	17.4	723	8	ADN9696 Novel hum
42	387.5	17.4	742	8	ABO58351 Human gen
43	387.5	17.4	839	7	ADJ70044 Human hea
44	387.5	17.4	872	7	ADC37535 Human nuc
45	386.5	17.3	327	8	ADR08458 Human pro

ALIGNMENTS

RESULT 1					
AA93316	standard; protein; 412 AA.				
ID	AA93316;				
XX					
AC	AA93316;				
XX					
DT	04-SEP-2000 (first entry)				
XX					
DE	A transcription factor designated CATP11A.				
XX					
KM	Transcription factor; CATP11A; DNA-binding protein;				
KW	ribosomal RNA 5S gene; fungal infection.				
XX					
OS	Candida albicans.				
XX					
FH	Key Location/Qualifiers				
FT	Misc-difference 193				
FT	/note= "Ser encoded by CTC"				
FT	Misc-difference 339				
FT	/note= "Ser encoded by CTC"				
XX					
PN	WO200028037-A1.				
XX					
PD	18-MAY-2000.				
XX					
PF	09-NOV-1999; 99WO-FR002739.				
XX					
PR	10-NOV-1998; 98FR-00014147.				
XX					
PA	(HMRI) HOECHST MARION ROUSSEL.				
XX					
PI	Bordon-Pallier F, Camier S, Sentenac A;				
XX					
DR	WPI; 2000-376549/32.				
DR	N-PSDB; AAA15398.				
XX					
PT	New nucleic acid encoding Candida albicans transcription factor, useful				
XX	e.g. in screening for antimycotic agents and for immunization.				
XX					
PS	Claim 12; Page 35-36; 45pp; French.				
XX					
CC	The present sequence represents a Candida albicans transcription factor,				
CC	designated CATP11A. The polypeptide is a DNA-binding protein, which is				
CC	involved in initiating transcription of the ribosomal RNA 5S gene. The				
CC	polynucleotide is used to screen for its specific inhibitors, potentially				
CC	useful as antimycotic agents, to raise an antibody response that is				
CC	protective against fungal infection and to raise antibodies. Such				
CC	antibodies, as well as the polypeptides and polynucleotides are used in				

CC Compositions for diagnosing and treating fungal infections, e.g. by
CC detecting polymorphisms and mutations
XX
XX
SQ Sequence 412 AA;
Query Match 100.0%; Score 2229; DB 3; Length 412;
Best Local Similarity 100.0%; Pred. No. 1e-167;
Matches 412; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MSESDETKSISLSISSSSSRPKKYYICTYEGCDKAYNRPSSLBQHLRTHSNDRPYKTYD 60
DB 1 MSESDETKSISLSISSSSSRPKKYYICTYEGCDKAYNRPSSLBQHLRTHSNDRPYKTYD 60
QY 61 DCDKAFPRKSHLETHIVSHSEKKPFCVCGKGVNSRQHLKRHEITHYKSPCTFENCQE 120
DB 61 DCDKAFPRKSHLETHIVSHSEKKPFCVCGKGVNSRQHLKRHEITHYKSPCTFENCQE 120
QY 121 AFYKHQSLRHHILSVHEKTLTKCKQCNKFTPRPSKLAQHLKHHGSPAYQCDHPGCFKNF 180
DB 121 AFYKHQSLRHHILSVHEKTLTKCKQCNKFTPRPSKLAQHLKHHGSPAYQCDHPGCFKNF 180
QY 181 QTWSVLQPHIKOSHPRKCPKCGKGVGKGLSSHMLSHDDSTMIKIWTCDYCDVGKFPK 240
DB 181 QTWSVLQPHIKOSHPRKCPKCGKGVGKGLSSHMLSHDDSTMIKIWTCDYCDVGKFPK 240
QY 241 KNELVENYNI FHDGNI PDLLKETEYVKLENLDOGSKLNNLHELETEKLYAEDEBDE 300
DB 241 KNELVENYNI FHDGNI PDLLKETEYVKLENLDOGSKLNNLHELETEKLYAEDEBDE 300
QY 301 DSLDEKSDVRSDSMSAQRISIKSFYASLBSKSVSLINSNGKINCPRNCCDRMFSREY 360
DB 301 DSLDEKSDVRSDSMSAQRISIKSFYASLBSKSVSLINSNGKINCPRNCCDRMFSREY 360
QY 361 DLRRHLKWHMDNLQRIEFLNSIEKEETPEGEPLVKKARMDLLPNETSVISR 412
DB 361 DLRRHLKWHMDNLQRIEFLNSIEKEETPEGEPLVKKARMDLLPNETSVISR 412
RESULT 2
ABP73656 standard; protein; 412 AA.
XX
XX
AC ABP73656;
XX
DT 30-JAN-2003 (first entry)
XX
DE Candida albicans essential protein SEQ ID NO 7493.
XX
KM Fungus; yeast; tetracyclin; promoter; GRACE strain; biosynthesis;
KM signal transduction; DNA replication; cell division; growth;
KM proliferation; Candida albicans; fungicide; antifungal.
XX
OS Candida albicans.
XX
PN WO200253728-A2.
XX
PD 11-JUL-2002.
XX
PF 26-DEC-2001; 2001WO-US049486.
XX
PR 29-DEC-2000; 2000US-0259128P.
PR 20-FEB-2001; 2001US-00792024.
PR 22-AUG-2001; 2001US-0314050P.
XX
PA (ELIT-) ELITRA PHARM INC.
XX
PI Roemer T, Jiang B, Boone C, Busey H, Ohlsen KU,
XX
DR WPI; 2002-566694/60.
DR N-PSDB; AB232206.
XX
PT Constructing strains for identifying gene products as effective targets
PT for therapeutic intervention, by inactivating in the strain one allele of

PT a gene and placing other allele of the gene under conditional expression.
XX
XX
XX Claim 44; SEQ ID NO 7493; 167pp + Sequence Listing; English.
CC The invention relates to constructing (M1) a strain of diploid fungal
CC cells in which both alleles of a gene are modified, comprising modifying
CC one allele by insertion or replacement by a cassette having an
CC expressible selectable marker and modifying other allele by
CC recombination, of a promoter replacement fragment with a heterologous
CC promoter, so that expression of the second allele is regulated by the
CC promoter. (M1) is useful for constructing a strain of diploid fungal
CC cells in which both alleles of a gene are modified. The diploid fungal
CC cells having both alleles modified are useful for identifying a gene that
CC is essential to the survival or growth of a fungus, a gene that
CC contributes to the virulence and/or pathogenicity of a fungus, a gene
CC that contributes to the resistance of a diploid fungus to an antifungal
CC agent, an antifungal agent that inhibits the growth of a diploid fungus
CC and for identifying a therapeutic agent for treatment of a mammalian
CC disease. (M1) is useful for identifying a compound which modulates the
CC activity of a gene product, preferably enzymatic activity, carbon
CC compound catabolism, biosynthetic, transporter, transcriptional,
CC translational, signal transduction, DNA replication and cell division
CC activity. The method is useful for identifying a compound having the
CC ability to inhibit growth or proliferation of C. albicans cells and for
CC treating infection by C. albicans. The present sequence is that of an
CC essential Candida albicans protein used in the method of the invention.
CC Note: The sequence data for this patent is not represented in the printed
CC specification but is based on sequence information supplied to Derwent by
CC the European Patent Office
XX
SQ Sequence 412 AA;
Query Match 99.8%; Score 2225; DB 5; Length 412;
Best Local Similarity 99.8%; Pred. No. 2.2e-167;
Matches 411; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
QY 1 MSESDETKSISLSISSSSSRPKKYYICTYEGCDKAYNRPSSLBQHLRTHSNDRPYKTYD 60
DB 1 MSESDETKSISLSISSSSSRPKKYYICTYEGCDKAYNRPSSLBQHLRTHSNDRPYKTYD 60
QY 61 DCDKAFPRKSHLETHIVSHSEKKPFCVCGKGVNSRQHLKRHEITHYKSPCTFENCQE 120
DB 61 DCDKAFPRKSHLETHIVSHSEKKPFCVCGKGVNSRQHLKRHEITHYKSPCTFENCQE 120
QY 121 AFYKHQSLRHHILSVHEKTLTKCKQCNKFTPRPSKLAQHLKHHGSPAYQCDHPGCFKNF 180
DB 121 AFYKHQSLRHHILSVHEKTLTKCKQCNKFTPRPSKLAQHLKHHGSPAYQCDHPGCFKNF 180
QY 181 QTWSVLQPHIKOSHPRKCPKCGKGVGKGLSSHMLSHDDSTMIKIWTCDYCDVGKFPK 240
DB 181 QTWSVLQPHIKOSHPRKCPKCGKGVGKGLSSHMLSHDDSTMIKIWTCDYCDVGKFPK 240
QY 241 KNELVENYNI FHDGNI PDLLKETEYVKLENLDOGSKLNNLHELETEKLYAEDEBDE 300
DB 241 KNELVENYNI FHDGNI PDLLKETEYVKLENLDOGSKLNNLHELETEKLYAEDEBDE 300
QY 301 DSLDEKSDVRSDSMSAQRISIKSFYASLBSKSVSLINSNGKINCPRNCCDRMFSREY 360
DB 301 DSLDEKSDVRSDSMSAQRISIKSFYASLBSKSVSLINSNGKINCPRNCCDRMFSREY 360
QY 361 DLRRHLKWHMDNLQRIEFLNSIEKEETPEGEPLVKKARMDLLPNETSVISR 412
DB 361 DLRRHLKWHMDNLQRIEFLNSIEKEETPEGEPLVKKARMDLLPNETSVISR 412
RESULT 3
ABJ25655 standard; protein; 564 AA.
XX
XX
AC ABJ25655;
XX
DT 16-APR-2003 (first entry)
XX

XX Novel human protein sequence #325.
DE anti-inflammatory; dermatological; neuroprotective; immunomodulator;
XX antibacterial; virucide; antiparasitic; cytostatic; gene therapy;
KM vaccine; inflammatory; CNS; immune disorder; cancer; psoriasis; diabetes;
KM early aging; hormonal imbalance; ischemic heart disease;
KM ulcerative colitis.
XX Homo sapiens.
OS
XX WO2004038003-A2.
XX PD 06-MAY-2004.
XX PF 24-OCT-2003; 2003MO-US033947.
XX PR 25-OCT-2002; 2002US-0421061P.
PR 25-OCT-2002; 2002US-0421080P.
PR 25-OCT-2002; 2002US-0421552P.
PR 25-OCT-2002; 2002US-0421614P.
PR 30-OCT-2002; 2002US-0422177P.
PR 30-OCT-2002; 2002US-0422178P.
PR 15-NOV-2002; 2002US-0426355P.
PR 15-NOV-2002; 2002US-0426384P.
PR 15-NOV-2002; 2002US-0426394P.
PR 15-NOV-2002; 2002US-0426430P.
PR 15-NOV-2002; 2002US-0426916P.
PR 27-NOV-2002; 2002US-0429224P.
PR 27-NOV-2002; 2002US-0429275P.
PR 27-NOV-2002; 2002US-0429302P.
PR 27-NOV-2002; 2002US-0429366P.
PR 27-NOV-2002; 2002US-0429651P.
PR 04-DEC-2002; 2002US-0430645P.
PR 04-DEC-2002; 2002US-0430651P.
PR 04-DEC-2002; 2002US-0430657P.
PR 04-DEC-2002; 2002US-0430663P.
PR 04-DEC-2002; 2002US-0430668P.
PR 05-DEC-2002; 2002US-0430684P.
PR 05-DEC-2002; 2002US-0430937P.
PR 05-DEC-2002; 2002US-0430965P.
PR 12-DEC-2002; 2002US-0431458P.
PR 12-DEC-2002; 2002US-0433251P.
PR 12-DEC-2002; 2002US-0433500P.
PR 13-DEC-2002; 2002US-0433316P.
PR 13-DEC-2002; 2002US-0433318P.
PR 23-DEC-2002; 2002US-0436236P.
PR 03-JAN-2003; 2003US-0437914P.
PR 17-JAN-2003; 2003US-0440820P.
PR 17-JAN-2003; 2003US-0440821P.
PR 18-APR-2003; 2003US-0463780P.
PR 18-APR-2003; 2003US-0463788P.
PR 18-APR-2003; 2003US-0463789P.
PR 18-APR-2003; 2003US-0463792P.
PR 02-MAY-2003; 2003US-0467199P.
PR 02-MAY-2003; 2003US-0467201P.
PR 02-MAY-2003; 2003US-0467203P.
PR 02-MAY-2003; 2003US-0467204P.
PR 19-MAY-2003; 2003US-0471306P.
PR 22-MAY-2003; 2003US-0472420P.
PR 22-MAY-2003; 2003US-0472430P.
PR 09-JUN-2003; 2003US-0476609P.
PR 09-JUN-2003; 2003US-0476621P.
PR 09-JUN-2003; 2003US-0476632P.
PR 09-JUN-2003; 2003US-0476641P.
PR 08-JUL-2003; 2003US-0485217P.
PR 08-JUL-2003; 2003US-0485218P.
PR 08-JUL-2003; 2003US-0485223P.
PR 08-JUL-2003; 2003US-0485224P.
PR 08-JUL-2003; 2003US-0485325P.
PR 08-JUL-2003; 2003US-0485359P.
PR 14-JUL-2003; 2003US-0486446P.

PR 14-JUL-2003; 2003US-0486480P.
PR 15-JUL-2003; 2003US-0486891P.
PR 15-JUL-2003; 2003US-0486960P.
PR 08-AUG-2003; 2003US-0493341P.
PR 08-AUG-2003; 2003US-0493370P.
PR 08-AUG-2003; 2003US-0493573P.
PR 08-AUG-2003; 2003US-0493577P.
XX
XX (FIVE-) FIVE PRIME THERAPEUTICS INC.
PA
XX Williams LT, Chu K, Lee E, Hestir K, Beaurang PA, Behrens D;
PI Halenbeck RF, Kochakota S, Lin H, Linnemann T, Pierce K, Wang Y;
PI Wong JCP, Wu G, Zhang H, Zeng C;
XX
DR WPI; 2004-365511/34.
DR N-PSDB; ADN98725.
XX
XX New nucleic acid molecules, useful in preparing a composition for
PT treating or preventing e.g. inflammatory, CNS, bacterial or viral
PT disorders, cancer, psoriasis, diabetes, ischemic heart disease or
PT ulcerative colitis.
XX
PS Claim 14; SEQ ID NO 1109, 532pp; English.
XX
XX The invention relates to a nucleic acid molecule comprising a
CC polynucleotide sequence or its complement that encodes a polypeptide. The
CC nucleic acid is useful in preparing a composition for treating or
CC preventing inflammatory, CNS, immune, bacterial or viral disorder,
CC cancer, psoriasis, diabetes, early aging, hormonal imbalance, ischemic
CC heart disease or ulcerative colitis. This sequence corresponds to a
XX protein of the invention.
SQ Sequence 664 AA;
Query Match 18.3%; Score 407; DB 8; Length 664;
Best Local Similarity 32.6%; Pred. No. 2.9e-23;
Matches 115; Conservative 46; Mismatches 138; Indels 54; Gaps 16;
QY 23 KYITCTYBEGCDKAYNRPSSLEQHLRTHNSDRPKCTVDDCDKAFRKSHEHTIVGHSK 82
DB 298 KPYQC--KEDCKVFSRKSRYLERRRHTGEEKPKCY--CDKAFRHSLSLAQHVITHTRE 353
QY 83 KPFHSGVCGVGNVSRQLKRNHEITHT--KSFCTFENCOEAFYKHSL-RHHITSVHEKT 139
DB 354 KPYKNECGKTREBNALVHKTIHTGEEKPKYC--NECGKVPYQSNLARRHRLHTGEEKP 411
QY 140 LTCKQCNKVFTRPSKLAQHLKHHGSPAYQCDHPGCFNPFQTSVLAQFH--IKOSHPRKL 197
DB 412 YKCEKCDKVFSSRSHLERRRIHTGEEKP-YKC--KYCDKAFRHSLSLAQHVITHTGEEKPY 468
QY 198 KCPKCGKCGVCKGLSSHMLSHDDSTMIKIMTCIDYDVCK-PAKKNELVEHYNIPIFDGNI 256
DB 469 KNECGKTFVONSLSIWMHVIH---TGEKRYKNEC--KSFPHKSLSLVYHNHL-HTGEEK 522
QY 257 PDDLKETEYKLENLDDQSKLNNLHELETEKLYKEEDBEDBEDSLDEKRSVDRSDSMS 316
DB 523 P---YKNECGKVFQSQSLACHRRHTGSE-KPYKEBEDK----- 559
QY 317 AGRSIKFTASLEGSKSVSLISNSGKTIKNCPCNDCRMFSRYEDLRRHLKWH 369
DB 560 -----VFNIKSHLEIHRVHTGEEKPKYC--RVCDKAFGRDSYLAQHV 602
RESULT 7
AAU30831
ID AAU30831 standard; protein; 1230 AA.
XX
AC AAU30831;
XX
DT 18-DEC-2001 (first entry)
XX
DE Novel human secreted protein #1322.
XX

KW Human; vaccination; gene therapy; nutritional supplement;
 KW stem cell proliferation; haematopoiesis; nerve tissue regeneration;
 KW immune suppression; immune stimulation; anti-inflammatory; leukaemia.
 OS Homo sapiens.
 XX WO200179449-A2.
 EN
 XX 25-OCT-2001.
 PD
 XX 16-APR-2001; 2001WO-US008656.
 PF
 XX 18-APR-2000; 2000US-00552929.
 PR 26-JAN-2001; 2001US-00770160.
 XX
 PA (HYSB-) HYSBQ INC.
 XX
 PI Tang YT, Liu C, Dmanac RT;
 XX
 DR WPI; 2001-611725/70.
 PT
 PT Nucleic acids encoding a range of human polypeptides, useful in genetic
 PT vaccination, testing and therapy.
 PS
 PS Claim 20; Page 355; 765pp; English.
 XX
 CC The invention relates to novel human secreted polypeptides. The
 CC polypeptides and antibodies to the polypeptides are useful for
 CC determining the presence of or predisposition to a disease associated
 CC with altered levels of polypeptide. The polypeptides are also useful for
 CC identifying agents (agonists and antagonists) that bind to them. Cells
 CC expressing the proteins are useful for identifying a therapeutic agent
 CC for use in treatment of a pathology related to aberrant expression or
 CC physiological interactions of the polypeptide. Vectors comprising the
 CC nucleic acids encoding the polypeptides and cells genetically engineered
 CC to express them are also useful for producing the proteins. The proteins
 CC are useful in genetic vaccination, testing and therapy, and can be used
 CC as nutritional supplements. They may be used to increase stem cell
 CC proliferation; to regulate haematopoiesis; and in bone, cartilage, tendon
 CC and/or nerve tissue growth or regeneration; immune suppression and/or
 CC stimulation; as anti-inflammatory agents; and in treatment of leukaemias.
 CC AAU28510-AAU33304 represent the amino acid sequences of novel human
 CC secreted proteins of the invention
 XX
 XX
 SQ Sequence 1230 AA;
 Query Match 18.1%; Score 403; DB 4; Length 1230;
 Best Local Similarity 32.9%; Pred. No. 1,4e-22;
 Matches 118; Conservative 45; Mismatches 144; Indels 52; Gaps 17;
 QY 20 SRPKYICTYEGCDKAYNRPSSLLEOHLRTHSNDRPYKCTVDDCDKAFPRKSHLETHIVSH 79
 DB 891 TREKPYKCC--EEGKAFSPQPSHLTTKRNHTGEKPYKC--EEGKAFSPQSSLTTHKIIH 946
 QY 80 SEKKPFHSGVCCGVNSRQHLKHEITHT--KSFKCTFENCCEAFYKQSL-RHHILSVH 136
 DB 947 TGEKPYKCEGCKAFKRSSTLTGHEKITHTGEKPYKC--EEGKAFSPQSSLTTRHTMHNG 1004
 QY 137 EKTLTQCKNKVTPTRSKLAQHLKHHGSPAYQCHPGCFKPFQWTVLQPHIKSHPK 196
 DB 1005 EKPYPKCEGCKAFNRSKLTTHKIIHTGKRP-YKCE--CGKAFSSSTLNGH-KKIHTR 1060
 QY 197 LKCPKCGKGVGK-----KGLSHMLSHDSTMIIMTCDYCDVGVKFAKNEVLVHNYI 250
 DB 1061 EKPYPK-EGC-GKAFSPQSSLTGKRIH--TGEKPYKCGEC--GKAKESALTKHKI 1113
 QY 251 FHDGNIPTDLLKETEYKLENLIDQSKLNLIHELETKLVYEDDEDESDIDEKRSDV 310
 DB 1114 IHTGE-----KPYKCECKCAFNQSSILTNHKKIITPPIH-----T 1151
 QY 311 RSDNSAQSISKFTSLGSKSVSKLISNGKINCPCNCDRNFSSRYDLRRHLKWH 369
 DB 1152 REKPYKYECGKGSFNRSSTFTK--HKVHTGVGLYKC--EEGCKSPFWSALTRHKKIH 1206

RESULT 8
 ADM05916
 ID ADM05916 strand; protein; 404 AA.
 XX
 XX ADM05916;
 AC
 XX 20-MAY-2004 (first entry)
 DT
 XX Human protein of the invention SEQ ID NO:4601.
 DE
 KW human; gene therapy; diagnostic marker; pharmaceutical.
 XX
 OS Homo sapiens.
 XX
 PN EP147046-A1.
 XX
 PD 24-SEP-2003.
 XX
 PF 12-APR-2002; 2002EP-00008400.
 XX
 PR 22-MAR-2002; 2002JP-00137785.
 XX
 PA (RESA-) RES ASSOC BIOTECHNOLOGY.
 XX
 XX Isogai T, Sugiyama T, Otsuki T, Wakamatsu A, Sato H, Ishii S;
 PI Yamamoto J, Isono Y, Hio Y, Otsuka K, Nagai K, Irie R, Tamechika I;
 PI Seki N, Yoshikawa T, Otsuka M, Nagahari K, Maeuho Y;
 DR WPI; 2003-723558/69.
 DR N-Psdb; ADM03473.
 XX
 PT New polynucleotides and polypeptides are useful in gene therapy, for
 PT developing a diagnostic marker or medicines for regulating their
 PT expression and activity, or as a target of gene therapy.
 XX
 PS Claim 1; SEQ ID NO 4601, 305pp; English.
 XX
 CC The invention relates to a novel human polynucleotide and the encoded
 CC polypeptide. A polynucleotide of the invention may have a use in gene
 CC therapy. An oligonucleotide of the invention ADM06202-ADM06773 is useful
 CC as a primer for synthesizing the polynucleotide or as a probe for
 CC detecting the polynucleotide. The polynucleotides ADM01316-ADM03758 are
 CC useful in gene therapy, for developing a diagnostic marker or medicines
 CC for regulating their expression and activity, or as a target of gene
 CC therapy. The proteins ADM03759-ADM06201 encoded by the polynucleotides
 CC are useful as pharmaceutical agents. The present sequence represents a
 CC protein sequence of the invention.
 XX
 XX
 SQ Sequence 404 AA;
 Query Match 17.9%; Score 400; DB 7; Length 404;
 Best Local Similarity 33.9%; Pred. No. 5.2e-23;
 Matches 121; Conservative 39; Mismatches 163; Indels 34; Gaps 19;
 QY 23 KVIITVYGCDCAYNRPSSLLEOHLRTHSNDRPYKCTVDDCDKAFPRKSHLETHIVSHSEK 82
 DB 39 KPYKC--EEGKAFKRSSTLTGHEKITHTGEKPYKC--EEGKGSSTLTTHKAIHABE 94
 QY 83 KPFHSGVCCGVNSRQHLKHEITHT--KSFKCTFENCCEAFYKQSLRHHILSVH--EK 138
 DB 95 KPYKCEGCKASNSSSKLMHKRIHTGKRPKC--EEGKAFSPQSSLTGHE-KRIHABEK 151
 QY 139 TLTQCKNKVTPTRSKLAQHLKHHGSPAYQCHPGCFKPFQWTVLQPH--IKQSHPK 196
 DB 152 PYKCEGCKAFNRSSTLTGHEKITHTGKRP-YKCE--GGCKAFSKVSTLNTKAIHABEKP 208
 QY 197 LKCPKCGKGVGKGLSHMLSHDSTMIIMTCDYCDVGVK-FAKNEVLVHNYI FHDGN 255
 DB 209 YKCEGCKASNSSSKLMHKRIH--TGEKPYKCEC--GKAFSPQSSLTGHEKRI-HAGE 262
 QY 256 IPDDLKETEYKLENLIDQSKLNLIHELETKLVYEDDEDESDIDEKRSDVRSDS 314

Db 263 KP---YKCECGCAFTWSSSFTGKRIHAE--KPYCECGKGFSTFSLTGHKIHTGE 318
 QY 315 --MSAQRSTKSFASLEGSKSVKLSNGKINCPCNNCDRMFSREYDLRHLKWH 369
 Db 319 KMYCECGKAF--SWSSILTEHKIHTGEKPYKC--EECGKAFSSSLTHHKRIH 371

RESULT 9

ADN06034
 ID ADN06034 standard; protein; 803 AA.

AC ADN06034;

DT 01-UTL-2004 (first entry)

DE Antipsoriatic protein sequence #1173.

KW antipsoriatic; gene therapy; psoriasis; diagnosis.

OS Homo sapiens.

PN WO2004028479-A2.

PD 08-APR-2004.

PF 25-SEP-2003; 2003WO-US030907.

PR 25-SEP-2002; 2002US-0414006P.

PA (GETH) GENENTECH INC.

PI Bodary S, Clark H, Jackman J, Schoenfeld J, Williams PM, Wood WI;

PI Wu TD;

DR WPI; 2004-305105/28.

DR N-PSDB; ADN06033.

PT New PRO nucleic acid or polypeptide, useful for preparing a
 PT pharmaceutical composition for diagnosing or treating psoriasis in a
 XX mammal.

PS Claim 9; SEQ ID NO 2429; 3069pp; English.

CC The invention relates to novel polynucleotide and polypeptides for
 CC treating psoriasis or a sequence having at least 80% identity to the
 CC above sequences. The nucleic acid is useful for preparing a composition
 CC for diagnosing or treating psoriasis in a mammal. This sequence
 CC corresponds to one of the polypeptides of the invention.

XX Sequence 803 AA;

Query Match 17.9%; Score 400; DB 8; Length 803;

Best Local Similarity 31.9%; Pred. No. 1,3e-22;

Matches 121; Conservative 53; Mismatches 171; Indels 34; Gaps 19;

QY 2 SESDETKISLSSSS--SRPKYICTYEGCDKAYNPSLLEQLRTSHNDPRPYKCTV 59

Db 367 TEGEAFRSNSNLTGKHITGEKPYKC--EECGKAFMSKSLTEHKLTHGKPYKC-- 422

QY 60 DDCDKAFKSHLETHIVSHSEKKPFHCSYCGKGVNSROLKHEIHT--KSFKCTFEN 117

Db 423 EECGKAFNWPSTLTGKHRIHTGEKPYCEVCGKAFNPSNLTGKHRIHTGEKPYKC--EE 480

QY 118 COEAFYGHOSL-RHILSVHKTTLTCQCNKVFTFRPEKLAQHKLKHGSPAYOCDHGC 176

Db 481 CGKAFSSSNLTGKHRIHTGEKPYCEVCGKAFNPSNLTGKHRIHTGEKPYKC--YKCEB--C 537

QY 177 FKQFQTVSLQFH--IKQSHPKLKCPCGKGVGKGLSSHMLSHDSTWIKIWTCDYCD 234

Db 538 GKAFNHSILTKKHRIHTGEKPYCEVCGKAFNPSNLTGKHRIHTGEKPYKC--TGEKPYKCEB-- 593

QY 235 VGR-FALKNELVEHYNI-FHGNIPDDLKTEYVKKLEMLDQSKANNLHLEPEKLVKE 293

Db 594 -GKAFQSSNLTGKHRIHTGEKPYKC--YKCECGKAFNPSNLTGKHRIHTGEKPYKC 647
 QY 294 E-DEDEDDSLDEKSDVRSDS--MSAQRSTKSFASLEGSKSVKLSNGKINCPCNNCDRMFSREYDLRHLKWH 369
 Db 648 ECGKAFNWPSTLTGKHRIHTGEKPYCEVCGKAFNPSNLTGKHRIHTGEKPYKC--YKCEB--C 537
 QY 351 NCDRMFSREYDLRHLKWH 369
 Db 705 -CGKAFNPSNLTGKHRIHTGEKPYKC--TGEKPYKCEB-- 593

RESULT 10

AA038689
 ID AA038689 standard; protein; 809 AA.

AC AA038689;

DT 22-OCT-2001 (first entry)

DE Human polypeptide SEQ ID NO 1834.

KW Human; nootropic; immunosuppressant; cytostatic; gene therapy; cancer;
 KW peripheral nervous system; neuropathy; central nervous system; CNS;
 KW Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
 KW amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;
 KW chemokine; thrombolytic; drug screening; arthritis; inflammation;
 KW leukaemia.

OS Homo sapiens.

PN WO200153312-A1.

PD 26-UTL-2001.

PF 26-DEC-2000; 2000WO-US034263.

PR 23-DEC-1999; 99US-00471275.

PR 21-JAN-2000; 2000US-00488725.

PR 25-APR-2000; 2000US-00552317.

PR 19-JUN-2000; 2000US-00598042.

PR 20-JUL-2000; 2000US-00620312.

PR 03-AUG-2000; 2000US-00653450.

PR 14-SEP-2000; 2000US-00662191.

PR 19-OCT-2000; 2000US-00693036.

PR 29-NOV-2000; 2000US-00727344.

PA (HYSR-) HYSEQ INC.

PI Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;
 PI Wang J, Wang Z, Weinman T, Xu C, Xue AJ, Yang Y, Zhang J, Zhao QH;
 PI Zhou P, Goodrich R, Drmanac RT;

DR WPI; 2001-442253/47.

DR N-PSDB; AA157845.

PT Novel nucleic acids and polypeptides, useful for treating disorders such
 PT as central nervous system injuries.

XX Example 3; SEQ ID NO 1834; 10078pp; English.

CC The invention relates to human nucleic acids (AA157798-AA161369) and the
 CC encoded polypeptides (AA138642-AA142213) with nootropic,
 CC immunosuppressant and cytostatic activity. The polynucleotides are useful
 CC in gene therapy. A composition containing a polypeptide or polynucleotide
 CC of the invention may be used to treat diseases of the peripheral nervous
 CC system, such as peripheral nervous injuries, peripheral neuropathy and
 CC localized neuropathies and central nervous system diseases, such as
 CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
 CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
 CC utilisation of the activities such as: Immune system suppression,
 CC Activin/inhibin activity, chemotactic/chemokine activity, haemostatic
 CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,

CC preventing inflammatory, CNS, immune, bacterial or viral disorder,
CC cancer, psoriasis, diabetes, early aging, hormonal imbalance, ischemic
CC heart disease or ulcerative colitis. This sequence corresponds to a
CC protein of the invention.

XX Sequence 444 AA;

Query Match 17.9%; Score 399; DB 8; Length 444;
Best Local Similarity 30.8%; Pred. No 7,1e-23;
Matches 120; Conservative 47; Mismatches 179; Indels 44; Gaps 18;

QY 20 SRPKYICTYEGCDKAYNRPSLLLEOHILRTSHNDPRPKCTVDDCDKAFPRKSHLETIVSH 79
Db 77 SGEKPYK--KECGKAVNBSANLSTHKRIHTGKKPYK--EECGKAFNRSLHTTKRIH 132
QY 80 SEKKPFPGCVGKGVNSRQHLKREHITHT--KSPKCTFENCQEAFFYHGL-RNHLISVH 136
Db 133 TGKKPYKCECGKAFNQSANLHTHKRIHTGKKPYK--EECGKAFNQSSTLTAKIKIHTG 190
QY 137 EKLTLTCKQCKKVFTRPSKLAQHKLKHGSGPACDHPGCFKNGQTSVLFPH--IKQSH 194
Db 191 EKPYKCECGKAFNQSSTLTTHKI-IHTGKPYKCEB--CGKAFSRSLHTTKRIHSGE 247
QY 195 PKLKCPKCGKGVGKGLSSHMLSHDSTMIKIWTCDYCDVGKFAKKEIYEHYNIFFHDG 254
Db 248 KPYKCECGKAFNQSSTLTTHKRIHAGE---KPYKCEVCSKA-PSRSHLTTHKRI-HTG 302
QY 255 NIPDDLKETBVKLENLDDGSKLNNLHELETEKLVEE-DEDEEDSLDEKRSQVRS 313
Db 303 EKP---YKCECGKAFNLSQTLTHKRIHTGE-KPYKCECGKAFNQSSTLSKRIYHTG 358
QY 314 S--MSAQRISIKSPFASLEGSKSVSKLISNGSKINCPCNNCDHPSREYDLRHLKMHND- 370
Db 359 EKPYKCECGKAFNQS--SHLTTHKRIHTGKPYK--EECGKAFNQSSTLTAKIKIHTG 414
QY 371 -----DNLQRISEFLNSIEKEE 387
Db 415 EKLKPESCNNACNIAKISKYKRNCAGEK 444

RESULT 12
ADN99690
ID ADN99690 standard; protein: 444 AA.
XX
AC ADN99690;
XX
DT 29-JUN-2004 (first entry)
XX
DE Novel human protein sequence #506.
XX
KW anti-inflammatory; dermatological; neuroprotective; immunomodulator;
KW antibacterial; virucide; antipsoriatic; cyostatic; gene therapy;
KW vaccine; inflammatory; CNS; immune disorder; cancer; psoriasis; diabetes;
KW early aging; hormonal imbalance; ischemic heart disease;
KW ulcerative colitis.
XX
OS Homo sapiens.
XX
PN WO2004038003-A2.
XX
PD 06-MAY-2004.
XX
PF 24-OCT-2003; 2003WO-US033947.
XX
PR 25-OCT-2002; 2002US-0421061P.
PR 25-OCT-2002; 2002US-0421080P.
PR 25-OCT-2002; 2002US-0421552P.
PR 25-OCT-2002; 2002US-0421614P.
PR 30-OCT-2002; 2002US-0422177P.
PR 30-OCT-2002; 2002US-0422178P.
PR 15-NOV-2002; 2002US-0426355P.
PR 15-NOV-2002; 2002US-0426384P.
PR 15-NOV-2002; 2002US-0426394P.

PR 15-NOV-2002; 2002US-0426430P.
PR 15-NOV-2002; 2002US-0426916P.
PR 27-NOV-2002; 2002US-0429224P.
PR 27-NOV-2002; 2002US-0429275P.
PR 27-NOV-2002; 2002US-0429302P.
PR 27-NOV-2002; 2002US-0429326P.
PR 27-NOV-2002; 2002US-0429651P.
PR 04-DEC-2002; 2002US-0430645P.
PR 04-DEC-2002; 2002US-0430651P.
PR 04-DEC-2002; 2002US-0430657P.
PR 04-DEC-2002; 2002US-0430663P.
PR 04-DEC-2002; 2002US-0430688P.
PR 04-DEC-2002; 2002US-0430684P.
PR 05-DEC-2002; 2002US-0430937P.
PR 05-DEC-2002; 2002US-0430965P.
PR 05-DEC-2002; 2002US-0431458P.
PR 12-DEC-2002; 2002US-0433251P.
PR 12-DEC-2002; 2002US-0433500P.
PR 13-DEC-2002; 2002US-0433318P.
PR 23-DEC-2002; 2002US-0436236P.
PR 03-JAN-2003; 2003US-0437914P.
PR 17-JAN-2003; 2003US-0440820P.
PR 17-JAN-2003; 2003US-0440821P.
PR 18-APR-2003; 2003US-0463700P.
PR 18-APR-2003; 2003US-0463708P.
PR 18-APR-2003; 2003US-0463716P.
PR 18-APR-2003; 2003US-0463732P.
PR 02-MAY-2003; 2003US-0467199P.
PR 02-MAY-2003; 2003US-0467201P.
PR 02-MAY-2003; 2003US-0467203P.
PR 02-MAY-2003; 2003US-0467230P.
PR 19-MAY-2003; 2003US-0471306P.
PR 19-MAY-2003; 2003US-047136P.
PR 22-MAY-2003; 2003US-0472420P.
PR 09-JUN-2003; 2003US-0476609P.
PR 09-JUN-2003; 2003US-0476621P.
PR 09-JUN-2003; 2003US-0476632P.
PR 09-JUN-2003; 2003US-0476641P.
PR 08-JUL-2003; 2003US-0485217P.
PR 08-JUL-2003; 2003US-0485218P.
PR 08-JUL-2003; 2003US-0485223P.
PR 08-JUL-2003; 2003US-0485224P.
PR 08-JUL-2003; 2003US-0485225P.
PR 08-JUL-2003; 2003US-0485325P.
PR 14-JUL-2003; 2003US-0485359P.
PR 14-JUL-2003; 2003US-0486446P.
PR 14-JUL-2003; 2003US-0486480P.
PR 15-JUL-2003; 2003US-0486891P.
PR 15-JUL-2003; 2003US-0486890P.
PR 08-AUG-2003; 2003US-0493341P.
PR 08-AUG-2003; 2003US-0493370P.
PR 08-AUG-2003; 2003US-0493573P.
PR 08-AUG-2003; 2003US-0493577P.

(FIVE-) FIVE PRIME THERAPEUTICS INC.

XX Williams LT, Chu K, Lee E, Healtix K, Beaurang PA, Behrens D,
PI Halenbeck RF, Kotchakota S, Lin H, Linemann T, Pierce K, Wang Y,
PI Wong JGP, Wu G, Zhang H, Zeng C;

DR WPI, 2004-365511/34.
DR N-PSDB; ADN98906.

XX New nucleic acid molecules, useful in preparing a composition for
PT treating or preventing e.g. inflammatory, CNS, bacterial or viral
PT disorders, cancer, psoriasis, diabetes, ischemic heart disease or
PT ulcerative colitis.

PS Claim 14; SEQ ID NO 1290; 532pp; English.

XX The invention relates to a nucleic acid molecule comprising a
CC polynucleotide sequence or its complement that encodes a polypeptide. The

CC nucleic acid is useful in preparing a composition for treating or
CC preventing inflammatory, CNS, immune, bacterial or viral disorder,
CC cancer, psoriasis, diabetes, early aging, hormonal imbalance, ischemic
CC heart disease or ulcerative colitis. This sequence corresponds to a
CC protein of the invention.

XX
XX Sequence 444 AA:

Query Match 17.9%; Score 398; DB 8; Length 444;
Best Local Similarity 30.2%; Pred. No. 8.5e-23;
Matches 121; Conservative 48; Mismatches 167; Indels 64; Gaps 18;

QY 20 SRPKYICTYEGCDYAYNRPSSLEOHILRTHSNDRPYKCTVDDCDKAFPRKSHLETHIVSH 79
DB 77 SGEKPYKC--KEGKAYNETSNLSTHKRIHTGKPYKC--EECGKAFNRLSHLTTHKRIIH 132
QY 80 SEKKPFHSGVCGVGNVSRHLLKHEITHT--KSPKCTENCQEAIFYKQSL--RHHLISVH 136
DB 133 TGGKPYKCEGKAFNOSANLTHHKRIHTGKPYKC--EECGKAFNOSSTLTTHKRIIHAG 190
QY 137 EKTLLTCKQCNKVFTRPSKLAOHILKHHGSPAYQCDHPCGFKNPOTWSVLQFPH--IKOSH 194
DB 191 EKPYKCEGKAFNOSSTLTTHKRI--HTGKPYKC--CGKAFNOSSTLTTHKRIHSGE 247
QY 195 PKLKCPKCGKCGVGGKGLSHMLSHDSTMIKWTCDYGVGKPAKNELVEHYNIPIHG 254
DB 248 KPYKCEGKAFNOSSTLTTHKRIHAGE--KPYKCEVCSKA--FSRFSHLTHKRI--HTG 302
QY 255 NIPDDLKETEYKVKLENLIDGSKLNNHLETEKLVKVEDDEDESDLSDEKSDVRS 314
DB 303 EKP---YKCEGKAFNLSQLTTHKRIHTG--KPYKCEC-----GKAFNOSST 348
QY 315 MSKORSI-----KSFASLESGSKSVSKLINSKKINCPKNNCDRMPFSREYD 361
DB 349 LSKHKYIHTGKPYKYKECGKAFNOS--SHLTTHKMIHTGKPYKC--EECGKAFNNSI 404
QY 362 LRRHLKMD-----DNLQRIEFLNSIEKEE 387
DB 405 LNRHKMIHTGKPYKYKECGKAFNOS--SHLTTHKMIHTGKPYKC--EECGKAFNNSI 444

RESULT 13

ID ABO60215 standard; protein; 487 AA.

AC ABO60215;

DT 29-JUL-2004 (first entry)

XX Human genome derived single exon protein #6449.

XX Human; gene expression; single exon probe; microarray;

KM alternative splicing event; genomic alteration.

XX Homo sapiens.

XX US2003194704-A1.

XX 16-OCT-2003.

XX 03-APR-2002; 2002US-00029386.

XX 03-APR-2002; 2002US-00029386.

XX (PENN/) PENN S G.

XX (HANK/) HANK D R.

XX (HANS/) HANZEL D K.

XX Penn SG, Rank DR, Hanzel DK,

XX WPI, 2004-119264/12.

PT New human genome-derived single exon nucleic acid probes useful for human

PT gene expression analysis, for identifying or characterizing alternative
PT splicing events, for assessing genomic alterations or as tools for
PT surveying tissues.

PS Claim 45; SEQ ID NO 33849; 80pp; English.

XX The invention relates to a nucleic acid probe for measuring human gene
XX expression, comprising any of the 27,400 fully defined nucleotide
XX sequences in the specification, or their complements or fragments, and
XX encoding at least 8 amino acids of any of the 6888 amino acid sequences
XX fully defined in the specification. The probe is a single exon probe that
XX hybridizes under high stringency conditions to a nucleic acid molecule
XX expressed in human cells or tissues. Also included are a spatially-
XX addressable set of single exon nucleic acid probes for measuring human
XX gene expression (comprising a plurality of single exon nucleic acid
XX probes cited above, where each of the plurality of probes is separately
XX and addressably isolatable or amplifiable from the plurality), a single
XX exon microarray for measuring human gene expression, a method of
XX measuring human gene expression, a vector comprising the single exon
XX probe cited above, an ORF-encoded peptide comprising at least 8
XX contiguous amino acids of any of the above-mentioned amino acid
XX sequences (optionally with conservative amino acid substitutions), an
XX isolated antibody that binds specifically to a peptide cited above,
XX methods of selling and/or licensing single exon probes or microarrays to
XX a customer desiring to measure gene expression, a method of providing
XX human gene expression data by subsequence, and a computer-readable
XX storage medium which contains a database having a plurality of records
XX (each record including data on the expression of a single exon probe
XX cited above. The probe, method and apparatus are useful in gene
XX expression analysis. The probes may be used as tools for surveying
XX tissues to detect the presence of expressed messages that contain their
XX specific exon, or in constructing genome-derived single exon microarrays.
XX In addition, the probes are used in identifying and characterizing
XX alternative splicing events, in detecting and characterizing gross
XX alterations in the genomic locus that includes their exon, in assessing
XX smaller genomic alterations, in priming the synthesis of nucleic acids,
XX or in expressing the ORF-encoded peptide. The present sequence is a human
XX single exon probe protein of the invention. Note: The sequence data for
XX this patent did not form part of the printed specification, but was
XX obtained in electronic format directly from USPTO at
XX seqdata.uspto.gov/sequence.html?docID=20030194704

XX Sequence 487 AA;

Query Match 17.9%; Score 398; DB 8; Length 487;
Best Local Similarity 30.2%; Pred. No. 9.7e-23;
Matches 121; Conservative 48; Mismatches 167; Indels 64; Gaps 18;

QY 20 SRPKYICTYEGCDYAYNRPSSLEOHILRTHSNDRPYKCTVDDCDKAFPRKSHLETHIVSH 79
DB 120 SGEKPYKC--KEGKAYNETSNLSTHKRIHTGKPYKC--EECGKAFNRLSHLTTHKRIIH 175
QY 80 SEKKPFHSGVCGVGNVSRHLLKHEITHT--KSPKCTENCQEAIFYKQSL--RHHLISVH 136
DB 176 TGGKPYKCEGKAFNOSANLTHHKRIHTGKPYKC--EECGKAFNRLSHLTTHKRIIHAG 233
QY 137 EKTLLTCKQCNKVFTRPSKLAOHILKHHGSPAYQCDHPCGFKNPOTWSVLQFPH--IKOSH 194
DB 234 EKPYKCEGKAFNOSSTLTTHKRI--HTGKPYKC--CGKAFNOSSTLTTHKRIHSGE 290
QY 195 PKLKCPKCGKCGVGGKGLSHMLSHDSTMIKWTCDYGVGKPAKNELVEHYNIPIHG 254
DB 291 KPYKCEGKAFNOSSTLTTHKRIHAGE--KPYKCEVCSKA--FSRFSHLTHKRI--HTG 345
QY 255 NIPDDLKETEYKVKLENLIDGSKLNNHLETEKLVKVEDDEDESDLSDEKSDVRS 314
DB 346 EKP---YKCEGKAFNLSQLTTHKRIHTG--KPYKCEC-----GKAFNOSST 391
QY 315 MSKORSI-----KSFASLESGSKSVSKLINSKKINCPKNNCDRMPFSREYD 361
DB 392 LSKHKYIHTGKPYKYKECGKAFNOS--SHLTTHKMIHTGKPYKC--EECGKAFNNSI 447
QY 362 LRRHLKMD-----DNLQRIEFLNSIEKEE 387

Db 448 LNRHMKIHTGEKLYKPESCNNACDNIATKISKYKRCAGKEK 487
 RESULT 14
 ABU96690
 ID ABU96690 standard; protein; 563 AA.
 XX
 AC ABU96690;
 XX
 DT 25-JUL-2003 (first entry)
 DE Human nucleic acid-associated protein (NAAP) #19.
 XX
 KW Human; nucleic acid-associated protein; cytosolic; antiarteriosclerotic;
 KW anticonvulsant; nootropic; neuroprotective; cerebroprotective; anti-HIV;
 KW antiatherogenic; antiinflammatory; chymotrypsin; gene therapy;
 KW cell proliferative disorder; cancer; atherosclerosis;
 KW neurological disorder; epilepsy; Huntington's disease; stroke;
 KW immune disorder; inflammatory disorder; AIDS; allergy;
 KW developmental disorder; Hypochyroidism; Cushing's syndrome; infection;
 KW protein-protein interaction; drug-target interaction;
 KW gene expression profile.
 XX
 OS Homo sapiens.
 XX
 PN MO2003023003-A2.
 XX
 PD 20-MAR-2003.
 XX
 PF 05-SEP-2002; 2002MO-US028540.
 XX
 PR 07-SEP-2001; 2001US-0317792P.
 PR 07-SEP-2001; 2001US-0317912P.
 PR 14-SEP-2001; 2001US-0322270P.
 PR 21-SEP-2001; 2001US-0324040P.
 PR 28-SEP-2001; 2001US-0326732P.
 PR 19-OCT-2001; 2001US-0346716P.
 PR 25-JAN-2002; 2002US-0351749P.
 PR 22-FEB-2002; 2002US-0359498P.
 XX
 PA (INCY-) INCYTE GENOMICS INC.
 XX
 PI Tang YT, Jackson JL, Griffin JA, Elliott VS, Forsythe TJ,
 PI Becha SD, Richardson TW, Lee EA, Sprague WM, Emerling BW,
 PI Thangavelu K, Warren BA, Tran UK, Yue H, Xu Y, Yue H, Li UX;
 PI Hafalia AJA, Sanjanwala B, Marquis JP, Gorvay AE, Lee SY, Ison CH;
 PI Baughn MR, Chawla NK, Nguyen DB, Swarnakar A, Zebardjian Y, Shah P,
 PI Thornton M, Yao MG, Khan FA, Gandhi AR, Yang J, Kable AE,
 PI Burford N, Rankumar J;
 XX
 DR WFI; 2003-313243/30.
 DR N-PSDB; ACA98938.
 XX
 PT New human nucleic acid associated proteins (NAAP), useful for diagnosing,
 PT treating and preventing diseases or conditions associated with the
 PT aberrant NAAP expression e.g. cancer, AIDS, atherosclerosis, epilepsy, or
 PT infections.
 XX
 PS Claim 1; Page 255-257; 345pp; English.
 XX
 CC The invention describes a novel human isolated nucleic acid-associated
 CC polypeptide (NAAP). The polypeptides and polynucleotides are useful in
 CC diagnosing, treating and preventing diseases or conditions associated
 CC with the decreased expression or overexpression of NAAP, such as cell
 CC proliferative (e.g. cancer, atherosclerosis), neurological (e.g.
 CC epilepsy, Huntington's disease, stroke), immune/inflammatory (e.g. AIDS,
 CC allergies) and developmental (e.g. Hypochyroidism, Cushing's syndrome)
 CC disorders, or infections. These are also useful in assessing the effects
 CC of exogenous compounds on the expression of nucleic acid and amino acid
 CC sequences of NAAP. The NAAP or its fragments are useful in screening
 CC compounds for effectiveness as agonist or antagonist of the polypeptides,
 CC or in altering the expression of the target polynucleotide and compounds

CC that specifically bind to or modulate the activity of the polypeptide.
 CC The microarray is useful in monitoring or measuring protein-protein
 CC interactions, drug-target interactions, and gene expression profiles.
 CC This is the amino acid sequence of a novel human nucleic acid-associated
 CC protein (NAAP)
 CC
 XX
 SQ Sequence 563 AA;
 Query Match 17.9%; Score 398; DB 6; Length 563;
 Best Local Similarity 30.2%; Pred. No. 1,2e-22;
 Matches 121; Conservative 48; Mismatches 167; Indels 64; Gaps 18;
 QY 20 SPPKYYICTEGCGDKAYNRPSSLLEOHILRTHSNDPRKYVDQCDKAFPRKSHLETHIVSH 79
 Db 196 SGEKPYKC--KECGKAYNTSNUSTHKRIHTGKPYKC--EECGKAFNNLSHLLTHKRIH 251
 QY 80 SEKKPFGVCGKGVNSRQHLRHEITHT--KSFCTFENCQEPAYKHQSL-RHHILSYH 136
 Db 252 TGKKPYKCECGKAFNQSANTLTHKRIHTGEKPYKC--EECGKAFNQSSTLTFAHKRIHAG 309
 QY 137 EKTITCKQCNKYFTRPSKLAQHKLKHNGSPAYQCDHPGCFNFTQWSVLQFH--IKQSH 194
 Db 310 EKPYPKCECGKAFNQSSTLTTHKI--HTGEKPYKCE--CGKAFNQSSTLTTHKRIHSGE 366
 QY 195 PKLKCPKCGKCGVKKGLSSHMLSHDSTMIKIWTCDYCDVGKFAKNELVEHYNIFFHQ 254
 Db 367 KPYKCECGKAFNQSSTLTTHKRIHAG--KPYKCEVSKA-FSFPSTLTHKRI--HTG 421
 QY 255 NIPDDLKTEYKLENNLDQSKLNNIHELETKLVEDEDEBDSLDEKSDVRS 314
 Db 422 EKP--YKCECGKAFNLSQTLTHKRIHTGE-KPYKCEC-----GKAFNQSST 467
 QY 315 MSAQSI-----KSFASLEGSKSVSKLINSKKINCPKNCNDRMFSREYD 361
 Db 468 LKHKVITHGEKPYKCEGKAFNOS--SHLTHKRIHTGEKPYKC--EECGKAFNNSI 523
 QY 362 LRRHLKMH-----DNLQRIESFNSIEKEE 387
 Db 524 LNRHMKIHTGEKLYKPESCNNACDNIATKISKYKRCAGKEK 563
 RESULT 15
 ADI60246
 ID ADI60246 standard; protein; 563 AA.
 XX
 AC ADI60246;
 XX
 DT 15-APR-2004 (first entry)
 XX
 DE Secreted polypeptide #130.
 XX
 KW osteopathic; vulnary; cytosolic; gene therapy; diagnosis; forensics;
 KW gene mapping; mutation identification; biodiversity; chromosome marker;
 KW immune response; myeloid cell disorder; lymphoid cell disorder;
 KW bone cartilage; tendon; ligament; nerve tissue growth; wound healing;
 KW burns; incision; ulcer; cancer.
 XX
 OS Homo sapiens.
 XX
 PN MO2003025142-A2.
 XX
 PD 27-MAR-2003.
 XX
 PF 18-SEP-2002; 2002MO-US029636.
 XX
 PR 18-SEP-2001; 2001US-0323349P.
 PR 16-SEP-2002; 2002US-00323349.
 XX
 PA (HYSE-) HYSEQ INC.
 XX
 PI Tang YT, Asundi V, Goodrich RW, Ren F, Zhang J, Zhao QA, Wang J;
 PI Ghosh M, Xue AJ, Wehrman T, Weng G, Zhou P, Dmanac RT;

DR WPI; 2003-354601/33.
DR N-PSDB; ADI60591.

DR N-PSDB; ADI60591.

PT New polyelectrolytes and secreted proteins, useful for treating myeloid or lymphoid cell disorders, in bone cartilage, tendon, ligament and nerve tissue growth or regeneration, in wound healing, and in tissue repair and replacement.

PS Claim 20; SEQ ID NO 281; 243pp; English.

The invention relates to novel isolated polynucleotides or a sequence encoding a polypeptide with biological activity, where the polynucleotide hybridizes to the polynucleotide under stringent hybridization conditions or has greater than 99% sequence identity with the polynucleotide. The polynucleotides and polypeptides are useful in diagnostics, forensics, gene mapping, identification of mutations responsible for genetic disorders and other traits, to assess biodiversity, as nutritional sources or supplements. The polynucleotides may also be used as molecular weight markers, chromosome markers or map related gene positions, or as an antigen to raise anti-DNA antibodies or elicit immune response. The polypeptides are useful for raising antibodies, as markers for tissues in which the corresponding polypeptide is expressed, for re-engineering damaged or diseased tissues, for treating myeloid or lymphoid cell disorders, in bone cartilage, tendon, ligament and/or nerve tissue growth or regeneration, in wound healing, in tissue repair and replacement, in healing of burns, incisions and ulcers, and in treating cancer. This sequence corresponds to a protein sequence of the invention.

SQ Sequence 563 AA;

Query Match	17.9%	Score 398	DB 7	Length 563
Best Local Similarity	30.2%	Pred. No. 1.2e-22		
Matches 121; Conservative	48	Mismatches 167	Indels 64	Gaps 18

Qy	20	BARPKYICTYBECDCAXNRPBLLBQHRNHSNDRPYKCTVVDODCKAPPRKSHLETHVSH	79
Db	196	SGEKPYKC--KECGAIVNETSNLSTHKRIHTGKRPKC--EECGAFNLSHLTTHKRIH	251
Qy	80	SEKPFHCSVCGVNSRQHLRHRHETHT--KSPFKTFENCBOAFYKHOSL-RHHLISYA	136
Db	252	TGKRPYKBECCGKANQSANLTHKRIHTGKRPYC--BECGAFQSSTLTTHKRIHAG	309
Qy	137	EKTLTCKOANKVTPRPSKLAQHKLNHGGSPAYQCDHPGCFXNPFQTVSVLOFH-IGOSH	194
Db	310	EKPVCCECGKAFQSSTLTTHKRI-INTGEKPYKCEB--CGKAFQSLSHLTTHKRIHSGE	366
Qy	195	PKLKPCKGKGCVGKKGLSSNHLNHDSDSMITKIMTCYDVCYKFAQGNELVEHYNI.FHOG	254
Db	367	KPYKBECCGKAFQSSSTLTTHKRIHAGE--KPYKEVCVSKA-FSPFSLTTHKRI-HYG	421
Qy	255	NIPDOLLKETEYKLENLIDQSGKLNNLHLELETEKLVAEDEBDEBDSLDEKRSVDRSPS	314
Db	422	EKP---YKCEBGKAFNLSQLTTHKRIHTGE-KPYKCEC-----GKAFQSSLT	467
Qy	315	MSAORSI-----KSFASLEBGSKVSKLISNSGKKINCPKONCDPMFSEYD	361
Db	468	LSKHVITHTGEKPYKEBCKAFNQS--SHLTTHKRIHTGKRPYC--BECGAFNNSI	523
Qy	362	LRRHLKMH-----DNLQRLESTLNSIEKEE	387
Db	524	LNRHMHITGEXLYKPBSCNMAADIAKISKYKRNCAAGK	563

Search completed: February 8, 2005, 15:18:59
Job time : 79 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 8, 2005, 15:14:56 ; Search time 42 Seconds
(without alignments)
732.272 Million cell updates/sec

Title: US-09-831-804-3

Perfect score: 2229
Sequence: 1 MSSEDTKSISSLSISSSSS.....PLVKARMDLPNETSVISR 412

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 513545 seqs, 74649064 residues

Total number of hits satisfying chosen parameters: 513545

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents AA:*
1: /cgn2_6/ptodata/1/1aa/5A.COMB.pep:*
2: /cgn2_6/ptodata/1/1aa/5B.COMB.pep:*
3: /cgn2_6/ptodata/1/1aa/6A.COMB.pep:*
4: /cgn2_6/ptodata/1/1aa/6B.COMB.pep:*
5: /cgn2_6/ptodata/1/1aa/PCUS.COMB.pep:*
6: /cgn2_6/ptodata/1/1aa/backfile1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1036	46.5	208	4 US-09-248-796A-19287	Sequence 19287, A
2	671	30.1	153	4 US-09-248-796A-19289	Sequence 19289, A
3	400	17.9	803	4 US-09-538-092-1026	Sequence 1026, Ap
4	399.5	17.9	656	4 US-09-949-016-10075	Sequence 10075, A
5	395.5	17.7	423	1 US-08-523-376-3	Sequence 3, Appl1
6	393.5	17.7	1191	4 US-09-949-016-6356	Sequence 6356, Ap
7	358	16.1	457	4 US-09-389-956-68	Sequence 68, Appl1
8	358	16.1	647	4 US-09-389-956-6	Sequence 6, Appl1
9	351	15.7	487	4 US-09-949-016-10055	Sequence 10055, A
10	350	15.7	685	4 US-09-949-016-8627	Sequence 8627, Ap
11	344	15.4	633	4 US-09-949-016-9434	Sequence 9434, Ap
12	343	15.4	746	4 US-09-949-016-11494	Sequence 11494, A
13	343	15.4	769	4 US-09-949-016-8149	Sequence 8149, Ap
14	343	15.4	810	4 US-09-538-092-11275	Sequence 11275, Ap
15	337	15.1	469	4 US-09-538-092-1195	Sequence 1195, Ap
16	333	14.9	642	4 US-09-949-016-6986	Sequence 6986, Ap
17	332	14.9	589	4 US-09-949-016-10374	Sequence 10374, A
18	329	14.8	589	4 US-09-538-092-1194	Sequence 1194, Ap
19	329	14.8	728	2 US-08-475-844-5	Sequence 5, Appl1
20	329	14.8	728	2 PCT-US95-08429-5	Sequence 5, Appl1
21	322	14.4	727	2 US-08-475-844-9	Sequence 9, Appl1
22	322	14.4	727	5 PCT-US95-08429-9	Sequence 9, Appl1
23	322	14.4	736	4 US-09-949-016-9241	Sequence 9241, Ap
24	321	14.4	654	4 US-09-538-092-1193	Sequence 1193, Ap
25	321	14.4	654	4 US-09-949-016-6357	Sequence 6357, Ap
26	321	14.4	663	4 US-09-949-016-7484	Sequence 7484, Ap
27	319.5	14.3	711	2 US-08-820-170A-10	Sequence 10, Appl1

28	319.5	14.3	711	3 US-09-055-699-10	Sequence 10, Appl1
29	319.5	14.3	711	3 US-09-273-565-10	Sequence 10, Appl1
30	319.5	14.3	711	3 US-09-565-538-10	Sequence 10, Appl1
31	319.5	14.3	711	3 US-09-661-468-10	Sequence 10, Appl1
32	319.5	14.3	711	4 US-09-976-165-10	Sequence 10, Appl1
33	319.5	14.3	711	4 US-09-828-648-2	Sequence 1356, Ap
34	317	14.2	321	4 US-09-538-092-11304	Sequence 1304, Ap
35	316.5	14.2	321	4 US-09-949-016-6364	Sequence 6364, Ap
36	316.5	14.2	321	4 US-08-933-750C-17	Sequence 17, Appl1
37	316	14.2	488	3 US-09-234-613-17	Sequence 17, Appl1
38	316	14.2	488	3 US-09-881-578A-4	Sequence 4, Appl1
39	315.5	14.2	314	4 US-09-949-016-8516	Sequence 8516, Ap
40	315.5	14.1	803	3 US-09-063-035-2	Sequence 2, Appl1
41	315	14.1	532	4 US-09-389-956-80	Sequence 80, Appl1
42	314.5	14.1	506	4 US-09-538-092-11177	Sequence 1177, Ap
43	313	14.0	506	4 US-09-881-578A-2	Sequence 2, Appl1
44	313	14.0	717	4 US-09-362-123A-4	Sequence 4, Appl1
45	312.5	14.0	543	4 US-09-362-123A-4	Sequence 4, Appl1

ALIGNMENTS

```
RESULT 1
US-09-248-796A-19287
; Sequence 19287, Application US/09248796A
; Patent No. 6747137
; GENERAL INFORMATION:
; APPLICANT: Kelch Weinstock et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBICA
; FILE REFERENCE: 107196.132
; CURRENT APPLICATION NUMBER: US/09/248,796A
; PRIOR FILING DATE: 1999-02-12
; PRIOR APPLICATION NUMBER: US 60/074,725
; PRIOR FILING DATE: 1998-02-13
; PRIOR APPLICATION NUMBER: US 60/096,409
; NUMBER OF SEQ ID NOS: 28208
; SEQ ID NO 19287
; LENGTH: 208
; TYPE: PRT
; ORGANISM: Candida albicans
US-09-248-796A-19287

Query Match          46.5%; Score 1036; DB 4; Length 208;
Best Local Similarity 96.2%; Pred. No. 3.8e-90;
Matches 202; Conservative 2; Mismatches 2; Indels 4; Gaps 2;

QY      204 KG-CVKKKGLSSHTLSDHDSMTKIKWTCDYCDVGRKAKKNEIYEHYNIFFDGNIPDDLK 262
          2 KGCLDKKKGLSSHTLSDHDTMKIKWTCDYCDVGRKAKKNEIYEHYNIFFDGNIPDDLK 61
          263 ETEVKKLENLDDGSKLNHLHETEKLVKEDEDEDESDLRKRSQVDSNSAQRISIK 322
          62 ETEVKKLENLDDGSKLNHLHETEKLVKEDEDEDESDLRKRSQVDSNSAQRISIK 118
QY      323 SFTASLEGSKSVKLSINSKKKINCPKNCNDRMFSREYDRLRLKMHDDNLQRIEFPINS 382
          119 SFTASLEGSKSVKLSINSKKKINCPKNCNDRMFSREYDRLRLKMHDDNLQRIEFPINS 178
QY      383 IEKEETPEGRPLVKARMDLPNETSVISR 412
          179 IEKEETPEGRPLVKARMDLPNETSVISR 208

RESULT 2
US-09-248-796A-19289
; Sequence 19289, Application US/09248796A
; Patent No. 6747137
; GENERAL INFORMATION:
; APPLICANT: Kelch Weinstock et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBICA
```

```

: TITLE OF INVENTION: FOR DIAGNOSTICS AND THERAPEUTICS
:
: FILE REFERENCE: 107196.132
:
: CURRENT APPLICATION NUMBER: US/09/248,796A
:
: CURRENT FILING DATE: 1999-02-12
:
: PRIOR APPLICATION NUMBER: US 60/074,725
:
: PRIOR FILING DATE: 1998-02-13
:
: PRIOR APPLICATION NUMBER: US 60/096,409
:
: PRIOR FILING DATE: 1998-08-13
:
: NUMBER OF SEQ ID NOS: 28208
:
: SEQ ID NO 19289
:
: LENGTH: 153
:
: TYPE: PRT
:
: ORGANISM: Candida albicans
:
: US-09-248-796A-19289

```

Query Match 30.1%; Score 671; DB 4; Length 153;
 Best Local Similarity 98.4%; Pred. No. 9.9e-56;
 Matches 122; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 MSSESDERKSTSLISSSSSSRPXYITCTYEGCDKXVYNPSSLLEOHLRTHSNDPRKYCTVD 60
DB 18 MSSEDERKSTSLISSSSSSRPKXICTYEGCDKXVYNPSSLLEOHLRTHSNDPRKYCTVE 77
QY 61 DCDKAPFRKSHLETHIVSHSEKKPHCVGCGVNSROHLXRHEITHTKSFCTFENCQE 120
DB 78 DCDKAPFRKSHLETHIVSHSEKKPHCVGCGVNSROHLXRHEITHTKSFCTFENCQE 137
QY 121 APYK 124
DB 138 TPYK 141

RESULT 3
US-09-538-092-1026
; Sequence 1026, Application US/09538092

```

? Patent NO. 6/55314
? GENERAL INFORMATION:
? APPLICANT: Glotc, Loic
? APPLICANT: Mansfield, Traci A.
? TITLE OF INVENTION: Protein-Protein Complexes and Method of Using Same
? FILE REFERENCE: 15966-542
? CURRENT APPLICATION NUMBER: US/09/538,092
? CURRENT FILING DATE: 2000-03-29
? PRIOR APPLICATION NUMBER: 60/127,352
? PRIOR FILING DATE: 1999-04-01
? PRIOR APPLICATION NUMBER: 60/178,965
? PRIOR FILING DATE: 2000-02-01
? NUMBER OF SEQ ID NOS: 1387
? SOFTWARE: Curatalseqformatter Version 0.9
? SEQ ID NO 1026
? LENGTH: 803
? TYPE: PRT
? ORGANISM: Homo sapiens
? FEATURE:
? NAME/KEY: misc_feature
? LOCATION: (0)...(0)
? OTHER INFORMATION: Polypeptide Accession Number P28160
? OS-09-538-092-1026

```

Query Match	17.9%	Score 400;	DB 4;	Length 803;
Best Local Similarity	31.9%	Pred. No. 5.4e-29;		
Matches 121; Conservative	53;	Mismatches 171;	Indels 34;	Gaps 19;

[illegible]

Db 481 CGAARSRSSNLTKHKKIHTEKKPYKCECGAFAKMSLTLPHKITHTEGKP--YKCE--C 537

Qy 177 FKNQFTWSVLQPH--IKSHPLKCKPKCGKCVGKGLSSHMLSHDSTWIKIWTCDYD 234

Db 538 GKAFNHFSLITLTKKIHTEGKPYKCECGKAFTOSSNLTKHKIH--TEGKPYKCEC-- 593

Qy 235 VGR-FAKKNELVEHNPIPHDGNIPDDLKETVEYKLTENLDQSKLNNHLELETKLVE 293

Db 594 -GAAPFOSSNLTKHKI-HTGSKP--YKCECGAFAQFSTLTKHKIH-TEKKPYKCE 647

Qy 294 E-DEDEDESLDEKTSVYRSDS--MSAQRISKSFASLEGSKSYSLISNGSKKINCPKN 350

Db 648 ECGKAPKMSSTLTGKKIHTEGKPYKCECGKAF--KLSSTLSYHKIITHTEGKPYKCEK- 704

Qy 351 NCDPMFSREYDRLRLKMH 369

Db 705 -CGKAFNRPSSNLIEHKKIH 722

```

RESULT 4
US-09-949-016-10075
; Sequence 10075, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FASTSEQ for Windows Version 4.0
; SEQ ID NO 10075
; LENGTH: 656
; TYPE: PRT
; ORGANISM: Human
US-09-949-016-10075

```

Query Match	17.9%;	Score 399.5;	DB 4;	Length 656;
Best Local Similarity	29.7%;	Pred. No. 4.5e-29;		
Matches 116;	Conservative 56;	Mismatches 158;	Indels 61;	Gaps 19

[illegible]


```

; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ. ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ. ID NO 8627
; LENGTH: 685
; TYPE: PRT
; ORGANISM: Human
US-09-949-016-8627

Query Match      15.4%; Score 350; DB 4; Length 685;
Best Local Similarity 28.8%; Pred. No. 2,4e-24;
Matches 110; Conservative 60; Mismatches 160; Indels 52; Gaps 19;

QY 6 ETKSISLSSSSSSPKKYICTYEGODKAYNRPILLLEOHLRTHSNDRPYKCTVDDCKA 65
DB 307 QNSLKKHQSHEKRP--YEC--NECGKAFRRSSNLIOHRIHSGEKPYCS--ECGKA 360
QY 66 FFKSHLETHIVSHSEKPPHSCGCVGNRSOHLKRHEITHT--KSFKCTFENCODAFY 123
DB 361 FRSSNLIKHRRHTHGEKPECECGKAFSGSAHLKHQVHTHGEKPYEC--NDCKGPRS 418
QY 124 KHOSL-RHHLIVHEKTLTKCKCNKVTPRPSKLAQHLKHHGSPAYQCDHPGCFKPNFOT 182
DB 419 RVSNLTKHHRVHTGKPKYKCDGKAFSSQSSSLIOHRIHTGKRP--HVCN--VCGKAFSY 475
QY 183 WSLYQPH--IKOSHPLKCPKCGKCGVKGKLSHMLSHDSTMIKITWTCDDYDVGK-PA 239
DB 476 SSVLRKHQIITHGEKPRCSVCGKAFSHSSALIIOHGVH--TGDPYACHEC--GTFG 530
QY 240 KKNELVEHYNIIFHDGNIIPDILLKETEVKLENLIDQSKLNLHLETEKLYEEDBEDE 299
DB 531 RSNLILHQRV-HTGEKPYCTCEGKTFSSSTLILQHQRHNL-----GLKHECNCQ- 581
QY 300 EDSLDEKRSQVRSDSMSAQRSI---KSFTASLEGSKVS-----KLISNGSKKINC 347
DB 582 -----GKAFNRSSNLILHOKVHTGKPYTC-VECGKFGSSSHLIOHRIHTGKRPYK 634
QY 348 PKNCCDMFSEYDLRHLKMH 369
DB 635 --SECGKAFSQRSLIOHRIH 654

RESULT 11
US-09-949-016-9434
; Sequence 9434, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; PRIOR FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ. ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ. ID NO 9434
; LENGTH: 639
; TYPE: PRT
; ORGANISM: Human
US-09-949-016-9434

Query Match      15.4%; Score 344; DB 4; Length 639;
Best Local Similarity 27.5%; Pred. No. 7.9e-24;
Matches 98; Conservative 47; Mismatches 129; Indels 82; Gaps 15;

QY 23 KYIYCTEGCDKAYNRPSSLLEOHLRTHSNDRPYKCTVDDCKAFRRSHLETHIVHSEK 82

```

```

DB 356 KPEYCIQ--CGKAHGHKALTDHLRIHTGEKPEYCA--ECGKTFRRSSNLIOHVSHTGE 411
QY 83 KPFHSCVCGKVNRSOHLKRHEITHTKSFCTFENQGEAFYHQSRLHLSVH--EKL 140
DB 412 KPEYCECKGSFRRYSSLTTEHVRTHGEIPEYCEKCGKAFKYSSTLTGK-MRIHTGEKRP 470
QY 141 TCKQCNKVTPRPSKLAQHLKHHGSPAYQCDHPGCFKPNFOTWSVLOPHIK--QSHPKLK 198
DB 471 ECGEKGKAFSKSHLIIHQRTHTKEK- YKCNB--CGKAFGSSSLTYMRTHTGSPPE 527
QY 199 CPKCGKGVGKGLSSHMLSHDSTMIKITWTCDDYDVGK-PAKKNELVEHYNIIFHDGNI 257
DB 528 CNOGCGKFRQIIBGLTQHQRVH--TGKPYECNKC--GKAFSGSHLIVHQRHT-FTG--- 578
QY 258 DDLKTEVKLENLIDQSKLNLHLETEKLYEEDBEDEDSLDEKRSQVRSDSMSA 317
DB 579 -----EKPYECNECEKAFNAKQQLVH 600
QY 318 ORSIKFTASLEGSKVSKLISNGK-INCPRNCCDMFSEYDLRHLKMHNDN 372
DB 601 QR-----SHTEKPYEC--NECGKTFKQNASLITKIVKTHSED 635

RESULT 12
US-09-949-016-11494
; Sequence 11494, Application US/09949016
; Patent No. 6812338
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; PRIOR FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ. ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ. ID NO 11494
; LENGTH: 746
; TYPE: PRT
; ORGANISM: Human
US-09-949-016-11494

Query Match      15.4%; Score 343; DB 4; Length 746;
Best Local Similarity 27.3%; Pred. No. 1.2e-23;
Matches 118; Conservative 65; Mismatches 141; Indels 108; Gaps 22;

QY 23 KYIYCTEGCDKAYNRPSSLLEOHLRTHSNDRPYKCTVDDCKAFRRSHLETHIVHSEK 82
DB 336 KPFKCT--ECGKAFRLSSKLIHQRIHTGKPYRC--EECGKAFGSSSLIHHQRHTGE 391
QY 83 KPFHSCVCGKVNRSOHLKRHEITHT--KSFCTFENCODAFYKHOSLNNH----- 131
DB 392 RPYGCREGCKAFSQSQSLVHQRHTHGERPYPC--KECGKAFQSSTLLOHQRHMTGEKA 449
QY 132 -ILS-----VH--EKLTKQCNKVTPRPSKLAQHLKHHGSPAYQCDHPG 176
DB 450 QILKASDSPSLVAHQHIAVERKPKDCBGKARWISRLSOHLITGKRP-YKCNK--C 506
QY 177 FKNFOTWSVLOPHIKOSH--PLKCPKCGKCGVKGKLSHMLSHD-STMIKITWTCY 232
DB 507 TKAFGSSSLIRH-QTHHTGKPEKFCDEGCKGFV-----QSHLIIHQRIHTGKRPVND 561
QY 233 CDVGK-PAKKNELVEHYNIIFHDGNIIPDILLKETEVKLENL-----DOGSKLNL 282
DB 562 C--GKAFSSSSSLIYHQRH-HKGEKPYECIQCCKAFSMTOLTIIHQVHTGKRPYCNKC 618

```

```
QY 283 -----HELETEKLVDEDE-----DEEDSLDEKSDVRSMSAQSISKF 324
DB 619 GKAFSONSTLFOHQIHAGVPECECGKAFSSSYLIEHQIHTRA-----QWPEYV 672
QY 325 TASLEGSYSKXISNSGKKI-----NCPK-----NCCRMES 357
DB 673 GNALEGSTVSRKKVNTIKKLHOCEDCEKIFRMRSLIHHRIHTEKPYKCDGKAFN 732
QY 358 REYDLRRHLKMH 369
DB 733 RSSRLTQHOKIH 744

RESULT 13
US-09-949-016-8149
; Sequence 8149, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; PRIOR FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 8149
; LENGTH: 769
; TYPE: PRT
; ORGANISM: Human
US-09-949-016-8149

Query Match 15.4%; Score 343; DB 4; Length 769;
Best Local Similarity 32.2%; Pred. No. 1.3e-23;
Matches 86; Conservative 35; Mismatches 100; Indels 46; Gaps 11;

QY 23 KKIYCTYEGCDKAYNPSSLLEOHLRTSHNDPRYKCTVDDCDKAFPRKSHLETHIVSHSR 82
DB 516 KPYKC--EECGKFSASFSQSHQVHTGKPRCNV--CGKGFSSQSYQAHQVHTG 571
QY 83 KPHKCVCGKGVNSRQHLKHEITHT--KSFCTFENCQDAFYKHOSLHHILSYH--EK 138
DB 572 KPYKCEVCGKRFVWSLNLNHHQRVHTGKPRYC--EECGKFSQASNLQAH--QSVHTG 628
QY 139 TLTKQCNKVFTRPSKLAQHLKHHGSPAYOCDDHGCFFNPQTWSVLQPH--IKQSHPK 196
DB 629 PRKCDACQKRFSSQASHLOAHQVHTGKPR--YKCD--TCGKAFSQRNLOVHQIHTGKPR 685
QY 197 LKCPKCGKGVCGKGLSHSLSHDSTMIKIWTCDYCDVQ----- 236
DB 686 FKCEBEGKESWAGSAGAHQRVH---TGKPYTCQCGKGFSGQASHFHTTHQVHTGERPY 742
QY 237 -----KFAKKNELVEHNIPIFDGNI 256
DB 743 ICDVCGKGFSSQRSLIYHQRVHTGNTL 769

RESULT 14
US-09-538-092-1275
; Sequence 1275, Application US/09538092
; Patent No. 6753314
; GENERAL INFORMATION:
; APPLICANT: Glot, Loic
; APPLICANT: Mansfield, Traci A.
; TITLE OF INVENTION: Protein-Protein Complexes and Method of Using Same
; FILE REFERENCE: 15966-542
; CURRENT APPLICATION NUMBER: US/09/538,092
```

```
QY 18 SSSRPKKYICTYEGCDKAYNPSSLLEOHLRTSHNDPRYKCTVDDCDKAFPRKSHLETHIV 77
DB 377 SHTGKPRFEC--NECGKAFSHKSLTLHQHTHTEKPYQNA--CGKTFCCQSDLTQKOR 432
QY 78 SHSEKPPHCSVCGKGVNSRQHLKHEITHT--KSFCTFENCQDAFYKHOSL--RHILLS 134
DB 433 THTGKLPYCEYCGKSPVNTSLKQVHTHTEKPRFEL--ECGKSFSEKSLTQHOKIH 490
QY 135 VHEKTLTKQCNKVFTRPSKLAQHLKHHGSPAYOCDDHGCFFNPQTWSVLQPHIKOSH 194
DB 491 IGDKSYECNACGKTLYHNSLTLRHQIHTGKPR--YECVE--CGKTFCCQSDLTQVH--QRTH 546
QY 195 PTLK---CPKCGKGVCGKGLSHSLSHDSTMIKIWTCDYCDVQK--PAKKNELVEHNI 250
DB 547 TGQKPRPACBCKSPFSKSTLSQHYRTH--TGKPYCEHC--SKITYNKSYLTQH--NR 600
QY 251 FHDGNIIPDDLKETEVEKLENLIDQSKYLNLIHELETEKLVKDEDE-----D 298
DB 601 THTGE-----KPYCNCGKAFYQKSL-----TQHRIHIEKPYKCNCGKARCH 647
QY 299 EEDSLDEKSDVRSMSAQSISFTASL-----EGSKV-----SKL 337
DB 648 KSAIIVHQRTHQEKPYKCNCGKSPCVKSGLIPIERKHTGKPYECNCGKFPYHKSLSL 707
QY 338 I-----SNSGKTKNCBKNNCDRMFSREYDLRRHLKMH 369
DB 708 TVHHRATHTEK--SCQCNCGKIPYKSELQAHQRSH 742

RESULT 15
US-09-538-092-1195
; Sequence 1195, Application US/09538092
; Patent No. 6753314
; GENERAL INFORMATION:
; APPLICANT: Glot, Loic
; APPLICANT: Mansfield, Traci A.
; TITLE OF INVENTION: Protein-Protein Complexes and Method of Using Same
; FILE REFERENCE: 15966-542
; CURRENT APPLICATION NUMBER: US/09/538,092
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: 60/127,352
; PRIOR FILING DATE: 1999-04-01
; PRIOR APPLICATION NUMBER: 60/178,965
; NUMBER OF SEQ ID NOS: 1387
; SOFTWARE: CuratPatSeqFormatter Version 0.9
; SEQ ID NO 1195
; LENGTH: 469
; TYPE: PRT
; ORGANISM: Homo sapiens
; NAME/KEY: misc_feature
; NAME/KEY: misc_feature

Query Match 15.4%; Score 343; DB 4; Length 810;
Best Local Similarity 28.3%; Pred. No. 1.4e-23;
Matches 112; Conservative 57; Mismatches 153; Indels 74; Gaps 20;
```

LOCATION: (0)...(0)
OTHER INFORMATION: Polypeptide Accession Number P52742
US-09-538-092-1195

Query Match 15.1%; Score 337; DB 4; Length 469;
Best Local Similarity 28.2%; Pred. No. 2,3e-23;

Matches 103; Conservative 49; Mismatches 163; Indels 50; Gaps 17;

QY	23	KKVICYEGGDKAYNRP	SLLEOH	LRTHSND	RPYKCTV	DDCKA	FFRKSH	LETHIV	SHSEK	82
DB	79	KPYKCTQ--CGRTFNQ	LAFLIQH	ORTHTG	EKPYECS--	EGCKSF	SFRSS	SPSQH	ERTHTGE	134
QY	83	KPFHCSV	CGKGVNSR	OH	LKREH	ITHT--	KSF	CTFENC	CEAFYKH	OSL--
DB	135	KPYECG	ECGKAF	QSFQ	SIHLTQ	HLRIHTG	EKPYQC--	GBCG	KAFSH	SSSLTKHQR
QY	140	LTCGQCN	KVFT	RP	SKLAQ	HKLKH	HGSPAY	QCDHP	GC	FKNFQ
DB	193	YECHEG	CKAFTQ	ITPLI	QHRTHTG	EKP--	YECGE--	CGKAF	SQSTLL	TEHRIHTG
QY	198	KCPKCG	KGC	VKGL	SGSHML	SHD	STM	IKITW	TC	DYGVK--
DB	250	GCNECG	KTF	SHSS	SLSQH	ERTH--	--TGEK	PECSQC--	GKAF	QSTHLTQ
QY	257	PDDLK	ETE	VYK	LEN	LDQ	SKLNN	LHE	LETE	KLYE
DB	303	-----	KPYEC	ND	CGKAF	SHSS	SLTKH	QRITHTG	EKPYEC	NC
QY	305	EKSDV	RS	DSMS	AQ	RSIK	SFT	ASL	EG	SKSV
DB	358	EKPYEC	NC	Q	CGRA	SARA	-----	TLLI	EHQR	-----
QY	365	HLKMH	369							
DB	407	HERTH	411							

Search completed: February 8, 2005, 15:23:40
Job time : 45 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Comugen Ltd.

OM protein - protein search, using sw model

Run on: February 8, 2005, 15:22:12 ; Search time 129 Seconds
(without alignments)
1040.228 Million cell updates/sec

Title: US-09-831-804-3

Perfect score: 2229

Sequence: 1 MESSDETKISLSLSSSSSS.....PLYKARMDLPNETVISR 412

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1373511 seqs, 325702437 residues

Total number of hits satisfying chosen parameters: 1373511

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Published Applications AA:*
1: /cgn2_6/prodata/2/pubppa/US07_PUBCOMB.pep:*
2: /cgn2_6/prodata/2/pubppa/US06_NEW_PUB.pep:*
3: /cgn2_6/prodata/2/pubppa/US06_PUBCOMB.pep:*
4: /cgn2_6/prodata/2/pubppa/US07_NEW_PUB.pep:*
5: /cgn2_6/prodata/2/pubppa/US07_PUBCOMB.pep:*
6: /cgn2_6/prodata/2/pubppa/PCTUS_PUBCOMB.pep:*
7: /cgn2_6/prodata/2/pubppa/US08_NEW_PUB.pep:*
8: /cgn2_6/prodata/2/pubppa/US08_PUBCOMB.pep:*
9: /cgn2_6/prodata/2/pubppa/US09A_PUBCOMB.pep:*
10: /cgn2_6/prodata/2/pubppa/US09B_PUBCOMB.pep:*
11: /cgn2_6/prodata/2/pubppa/US09C_PUBCOMB.pep:*
12: /cgn2_6/prodata/2/pubppa/US09_NEW_PUB.pep:*
13: /cgn2_6/prodata/2/pubppa/US10_PUBCOMB.pep:*
14: /cgn2_6/prodata/2/pubppa/US10_PUBCOMB.pep:*
15: /cgn2_6/prodata/2/pubppa/US10_PUBCOMB.pep:*
16: /cgn2_6/prodata/2/pubppa/US10_PUBCOMB.pep:*
17: /cgn2_6/prodata/2/pubppa/US10_NEW_PUB.pep:*
18: /cgn2_6/prodata/2/pubppa/US10_NEW_PUB.pep:*
19: /cgn2_6/prodata/2/pubppa/US60_NEW_PUB.pep:*
20: /cgn2_6/prodata/2/pubppa/US60_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2225	99.8	412	US-10-032-585-7493	Sequence 7493, Ap
2	445	20.0	564	US-10-128-714-3313	Sequence 3313, Ap
3	445	20.0	564	US-10-128-714-3313	Sequence 3313, Ap
4	400	17.9	404	US-10-108-260A-4601	Sequence 4601, Ap
5	398	17.9	487	US-10-029-386-33849	Sequence 33849, A
6	397.5	17.8	423	US-10-734-564-83	Sequence 83, Appl
7	395.5	17.7	409	US-10-106-698-5955	Sequence 5955, Ap
8	395.5	17.7	1191	US-10-408-765A-2105	Sequence 2105, Ap
9	392.5	17.6	595	US-10-203-0528-8	Sequence 8, Appl
10	392	17.6	555	US-10-094-749-2331	Sequence 2331, Ap
11	391	17.5	15	US-10-108-260A-3669	Sequence 3669, Ap
12	390	17.5	540	US-10-104-047-3748	Sequence 3748, Ap
13	390	17.5	576	US-10-094-749-2365	Sequence 2365, Ap

14	388	17.4	568	US-10-203-0528-7	Sequence 7, Appl
15	387.5	17.4	742	US-10-029-386-31985	Sequence 31985, A
16	387.5	17.4	839	US-10-408-765A-1850	Sequence 1850, Ap
17	385	17.3	577	US-10-381-327-3	Sequence 3, Appl
18	378	17.0	613	US-10-408-765A-1517	Sequence 1517, Ap
19	376	16.9	499	US-10-108-260A-2954	Sequence 2954, Ap
20	374	16.8	566	US-10-322-774-10	Sequence 10, Appl
21	373	16.7	452	US-10-471-450-5	Sequence 5, Appl
22	372.5	16.7	586	US-10-104-047-2592	Sequence 2592, Ap
23	372	16.7	393	US-09-911-261A-12	Sequence 12, Appl
24	372	16.7	393	US-10-057-408-12	Sequence 12, Appl
25	372	16.7	393	US-10-333-487-12	Sequence 12, Appl
26	371	16.6	403	US-10-029-386-32933	Sequence 32933, A
27	370	16.6	430	US-10-094-749-2723	Sequence 2723, Ap
28	369.5	16.6	564	US-10-108-260A-3590	Sequence 3590, Ap
29	367	16.5	514	US-10-029-386-34112	Sequence 34112, A
30	367	16.5	610	US-09-989-920-212	Sequence 212, App
31	367	16.5	610	US-10-363-616-421	Sequence 421, App
32	364.5	16.4	444	US-10-108-260A-3223	Sequence 3223, Ap
33	364	16.3	516	US-09-974-298-118	Sequence 118, App
34	364	16.3	535	US-10-264-048-3016	Sequence 3016, Ap
35	363	16.3	516	US-10-094-749-1976	Sequence 1976, Ap
36	361	16.2	803	US-10-408-765A-1049	Sequence 1049, Ap
37	360	16.2	529	US-10-108-260A-3085	Sequence 3085, Ap
38	360	16.2	791	US-10-471-450-18	Sequence 18, Appl
39	360	16.2	924	US-10-007-280A-231	Sequence 231, App
40	359.5	16.1	425	US-10-094-749-2297	Sequence 2297, Ap
41	359.5	16.1	427	US-10-755-889-370	Sequence 370, App
42	358.5	16.1	473	US-10-029-386-32047	Sequence 32047, A
43	358	16.1	406	US-09-764-864-976	Sequence 976, App
44	358	16.1	457	US-10-377-079-68	Sequence 68, Appl
45	358	16.1	647	US-10-377-079-6	Sequence 6, Appl

ALIGNMENTS

RESULT 1				
US-10-032-585-7493				
Sequence 7493, Application US/10032585				
Publication No. US20030180953A1				
GENERAL INFORMATION:				
APPLICANT: Terry, Roemer D.				
APPLICANT: Bo, Jjiang Boone				
APPLICANT: Charles, Boone				
TITLE OF INVENTION: Gene Disruption Methodologies for Drug Target Discovery				
FILE REFERENCE: 10182-005-999				
CURRENT APPLICATION NUMBER: US/10/032,585				
CURRENT FILING DATE: 2001-12-20				
NUMBER OF SEQ ID NOS: 8000				
SOFTWARE: PatentIn version 3.1				
SEQ ID NO 7493				
LENGTH: 412				
TYPE: PR				
ORGANISM: Candida albicans				
US-10-032-585-7493				
Query Match				
Best local Similarity 99.8%; Pred. No. 2,3e-154;				
Matches 411; Conservative 1; Mismatches 0; Indels 0; Gaps 0;				
Qy	1	MSSDETKISLSLSSSSSRPKKYCTYEGCDKAYNRPBLFOHLRTSHNDPRKYCTVD	60	
Db	1	MSSDETKISLSLSSSSSRPKKYCTYEGCDKAYNRPBLFOHLRTSHNDPRKYCTVD	60	
Qy	61	DCCKAFRRSHLTHVSHSEKRPFCVCGKVNROHKKHEITHTSFKTFENCOE	120	
Db	61	DCCKAFRRSHLTHVSHSEKRPFCVCGKVNROHKKHEITHTSFKTFENCOE	120	
Qy	121	AFYKHSLSRHLISVHEKTLTKCKCNKVFTRPSKLAQHLKHGSPAYOCDPGCGFKNF	180	
Db	121	AFYKHSLSRHLISVHEKTLTKCKCNKVFTRPSKLAQHLKHGSPAYOCDPGCGFKNF	180	

```
Qy 181 QWTVLQFHHKQSHPKLCKPCCKGCVGKKGSLSSHMLSHDDSTMIKWTCYCDVGVKFAK 240
| | | | |
Db 181 QWTVLQFHHKQSHPKLCKPCCKGCVGKKGSLSSHMLSHDDSTMIKWTCYCDVGVKFAK 240
Qy 241 KNELVENYNI FHDGNI PDDLKETEVEKYLENLLDOSKLNINHELETEKLYAEDEDEDE 300
| | | | |
Db 241 KNELVENYNI FHDGNI PDDLKETEVEKYLENLLDOSKLNINHELETEKLYAEDEDEDE 300
Qy 301 DSIDEKRSYVRSDSMSAORSIKSFTASLBSGSKVSXKLINSQKKNCPKNNCDRMFSREY 360
| | | | |
Db 301 DSIDEKRSYVRSDSMSAORSIKSFTASLBSGSKVSXKLINSQKKNCPKNNCDRMFSREY 360
Qy 361 DLRRHLKMHDDNIQRIESFLNLSIEKEETPEGEPLVYKAMDLLPNETSVISR 412
| | | | |
Db 361 DLRRHLKMHDDNIQRIESFLNLSIEKEETPEGEPLVYKAMDLLPNETSVISR 412
```

RESULT 2

```
US-10-128-714-3313
/ Sequence 3313, Application US/10128714
/ Publication No. US20030119013A1
/ GENERAL INFORMATION:
/ APPLICANT: Jiang, Bo
/ APPLICANT: Hu, Wengqi
/ APPLICANT: Tishkoff, Daniel
/ APPLICANT: Zamudio, Carlos
/ APPLICANT: Broshkin, Alexey M
/ APPLICANT: Lemieux, Sebastien M
/ TITLE OF INVENTION: Identification of Essential Genes in Aspergillus fumigatus and
/ FILE REFERENCE: 10182-018-999
/ CURRENT FILING DATE: 2002-04-23
/ PRIOR FILING DATE: 2002-04-23
/ PRIOR APPLICATION NUMBER: US 60/285,697
/ PRIOR FILING DATE: 2001-04-23
/ PRIOR APPLICATION NUMBER: US 60/287,066
/ PRIOR FILING DATE: 2001-04-27
/ PRIOR APPLICATION NUMBER: US 60/295,890
/ PRIOR FILING DATE: 2001-06-05
/ PRIOR APPLICATION NUMBER: US 60/303,899
/ PRIOR FILING DATE: 2001-07-09
/ PRIOR APPLICATION NUMBER: US 60/316,362
/ PRIOR FILING DATE: 2001-08-31
/ NUMBER OF SEQ ID NOS: 8603
/ SOFTWARE: PatentIn version 3.1
/ SEQ ID NO 3313
/ LENGTH: 564
/ TYPE: PRF
/ ORGANISM: Aspergillus fumigatus
US-10-128-714-3313
```

Query Match 20.0%; Score 445; DB 14; Length 564;

Best Local Similarity 30.3%; Pred. No. 3.9e-24;

Matches 125; Conservative 62; Mismatches 164; Indels 62; Gaps 21;

```
Qy 14 ISSSSSRP--KKYICTYEGCDKAYNRPSSLLEOHLRTSHNDPRPYKCTVDDCDKAFPRKS 70
| | | | |
Db 71 ISTSAKYPSELKTHRCPPFDGCTKAFNRPALQELHRSNNERIKCTCFEBCDKTFLRAS 130
Qy 71 HLETHIVS-HSEKKPFHCS--VCGKVNRSOHLKHEITH--TKSFKCT-FENCOZAPYK 124
| | | | |
Db 71 HLNHHIKSAHNGVDRDYCDRPGCGSFVTGSRRLRHLAHDGRDXYRCTEYPPCNETFRK 190
Qy 125 HOSLRHHILSVH--EKTLLTC-----KQCKNVFTRPSKLAQHLKHHGSPAYOCDH- 173
| | | | |
Db 125 HOSLRHHILSVH--EKTLLTC-----KQCKNVFTRPSKLAQHLKHHGSPAYOCDH- 173
Qy 191 HSTLQKIMTALHKKQRPQCPTHTPSTGQCTMAFDTAGHLRAHESRIH-TEKRSCTEC 249
| | | | |
Db 191 HSTLQKIMTALHKKQRPQCPTHTPSTGQCTMAFDTAGHLRAHESRIH-TEKRSCTEC 249
Qy 174 ----PGCFKNQFQWTVLQFHHKQSHPKLCKPCCKGCVGKKGSLSSHM-LSHDDSTM--IK 226
| | | | |
Db 174 ----PGCFKNQFQWTVLQFHHKQSHPKLCKPCCKGCVGKKGSLSSHM-LSHDDSTM--IK 226
Qy 250 SQHAGGAATFTPTVALLOAHIRSVHP-QCPNICALTCATSRELRLHLEVAHGDVSLBERK 308
| | | | |
Db 250 SQHAGGAATFTPTVALLOAHIRSVHP-QCPNICALTCATSRELRLHLEVAHGDVSLBERK 308
Qy 227 IWTCDY--CDVGKFAKNELVEHYNIFHDGNI-----PDDLKETEVEKYLENLLDQ---- 275
| | | | |
Db 227 IWTCDY--CDVGKFAKNELVEHYNIFHDGNI-----PDDLKETEVEKYLENLLDQ---- 275
```

```
Db 309 IFFCTVPGCD-RSFTKGNLTWHIRTVHQGEKRFVCGETDLSSSKKVSGMNN--DNGCGK 365
| | | | |
Qy 276 --GSKLNINHELETEKLYAEDEDEDEBDSL-----DEKRSYVRSDSMSAORSIKSFTASLE 329
| | | | |
Db 366 RYGSKLALAEHIRTALHLYQNAKARRORLGITTRROHSTATSPGVS-----LAALT 418
| | | | |
Qy 330 GSKSVSKLINSQKKNCPKNNCDRMFSREYDLRRHL--KMHDDNIQRIESFL 380
| | | | |
Db 419 GEGVAE-----TGRHIACTVESCPRHFRHDYDLWVMGSKHHFSEBETRDLFL 467
| | | | |
```

RESULT 3

```
US-10-128-714-8313
/ Sequence 8313, Application US/10128714
/ Publication No. US20030119013A1
/ GENERAL INFORMATION:
/ APPLICANT: Jiang, Bo
/ APPLICANT: Hu, Wengqi
/ APPLICANT: Tishkoff, Daniel
/ APPLICANT: Zamudio, Carlos
/ APPLICANT: Broshkin, Alexey M
/ APPLICANT: Lemieux, Sebastien M
/ TITLE OF INVENTION: Identification of Essential Genes in Aspergillus fumigatus and
/ FILE REFERENCE: 10182-018-999
/ CURRENT FILING DATE: 2002-04-23
/ PRIOR FILING DATE: 2002-04-23
/ PRIOR APPLICATION NUMBER: US 60/285,697
/ PRIOR FILING DATE: 2001-04-23
/ PRIOR APPLICATION NUMBER: US 60/287,066
/ PRIOR FILING DATE: 2001-04-27
/ PRIOR APPLICATION NUMBER: US 60/295,890
/ PRIOR FILING DATE: 2001-06-05
/ PRIOR APPLICATION NUMBER: US 60/303,899
/ PRIOR FILING DATE: 2001-07-09
/ PRIOR APPLICATION NUMBER: US 60/316,362
/ PRIOR FILING DATE: 2001-08-31
/ NUMBER OF SEQ ID NOS: 8603
/ SOFTWARE: PatentIn version 3.1
/ SEQ ID NO 8313
/ LENGTH: 564
/ TYPE: PRF
/ ORGANISM: Aspergillus fumigatus
US-10-128-714-8313
```

Query Match 20.0%; Score 445; DB 14; Length 564;

Best Local Similarity 30.3%; Pred. No. 3.9e-24;

Matches 125; Conservative 62; Mismatches 164; Indels 62; Gaps 21;

```
Qy 14 ISSSSSRP--KKYICTYEGCDKAYNRPSSLLEOHLRTSHNDPRPYKCTVDDCDKAFPRKS 70
| | | | |
Db 71 ISTSAKYPSELKTHRCPPFDGCTKAFNRPALQELHRSNNERIKCTCFEBCDKTFLRAS 130
Qy 71 HLETHIVS-HSEKKPFHCS--VCGKVNRSOHLKHEITH--TKSFKCT-FENCOZAPYK 124
| | | | |
Db 71 HLNHHIKSAHNGVDRDYCDRPGCGSFVTGSRRLRHLAHDGRDXYRCTEYPPCNETFRK 190
Qy 125 HOSLRHHILSVH--EKTLLTC-----KQCKNVFTRPSKLAQHLKHHGSPAYOCDH- 173
| | | | |
Db 125 HOSLRHHILSVH--EKTLLTC-----KQCKNVFTRPSKLAQHLKHHGSPAYOCDH- 173
Qy 191 HSTLQKIMTALHKKQRPQCPTHTPSTGQCTMAFDTAGHLRAHESRIH-TEKRSCTEC 249
| | | | |
Db 191 HSTLQKIMTALHKKQRPQCPTHTPSTGQCTMAFDTAGHLRAHESRIH-TEKRSCTEC 249
Qy 174 ----PGCFKNQFQWTVLQFHHKQSHPKLCKPCCKGCVGKKGSLSSHM-LSHDDSTM--IK 226
| | | | |
Db 174 ----PGCFKNQFQWTVLQFHHKQSHPKLCKPCCKGCVGKKGSLSSHM-LSHDDSTM--IK 226
Qy 250 SQHAGGAATFTPTVALLOAHIRSVHP-QCPNICALTCATSRELRLHLEVAHGDVSLBERK 308
| | | | |
Db 250 SQHAGGAATFTPTVALLOAHIRSVHP-QCPNICALTCATSRELRLHLEVAHGDVSLBERK 308
Qy 227 IWTCDY--CDVGKFAKNELVEHYNIFHDGNI-----PDDLKETEVEKYLENLLDQ---- 275
| | | | |
Db 227 IWTCDY--CDVGKFAKNELVEHYNIFHDGNI-----PDDLKETEVEKYLENLLDQ---- 275
Qy 276 --GSKLNINHELETEKLYAEDEDEDEBDSL-----DEKRSYVRSDSMSAORSIKSFTASLE 329
| | | | |
Db 276 --GSKLNINHELETEKLYAEDEDEDEBDSL-----DEKRSYVRSDSMSAORSIKSFTASLE 329
Qy 366 RYGSKLALAEHIRTALHLYQNAKARRORLGITTRROHSTATSPGVS-----LAALT 418
| | | | |
Db 366 RYGSKLALAEHIRTALHLYQNAKARRORLGITTRROHSTATSPGVS-----LAALT 418
```

QY 330 GSKSVSKLINSGKKINCRCNCCDMFSREYDLRRHL--KMHDDNTLQIESFL 380
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
Db 419 GEGYAE-----TGRHIACTVESCPHFPHRDYDLMVMSGKHHSSEERDRLF 467

```

RESULT 4
US-10-108-260A-4601
; Sequence 4601, Application US/10108260A
; Publication NO. US20040005560A1
; GENERAL INFORMATION:
; APPLICANT: HELIX RESEARCH INSTITUTE
; TITLE OF INVENTION: NO. US20040005560A1el full length cDNA
; FILE REFERENCE: H1-A01.06
; CURRENT APPLICATION NUMBER: US/10/108,260A
; CURRENT FILING DATE: 2002-03-27
; NUMBER OF SEQ ID NOS: 5458
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 4601
; LENGTH: 404
; TYPE: PRT
; ORGANISM: Homo sapiens
; US-10-108-260A-4601

```

Query Match	17.9%	Score 400	DB 15	Length 404
Best Local Similarity	33.9%	Pred. No. 5e-21		
Matches 121	Conservative 39	Mismatches 163	Indels 34	Gaps 19

Qy	83	KPHGVCVGJUNVRQHLKRHLLTHT--KSFECTENOGEPYKQSLRHLLHSVH--EK	138
Dp	39	KPKVC--EECGKAFNNSILITGHKIIHTGEKPYKC--EECGKGSFVSVTLNTHKAIHAE	94
Qy	23	KKIICTYEGGCDKAVNPSLLEOHLRHSNDRCYKTCVDDCKAFPRKSHLETHIVHSSEK	82
Dp	95	KPKCECGKAANSSSSKIMEHRKIHTGEKPYKC--EECGKAFSSSSLTHER-KRIHGEK	151
Qy	139	TLTKCOCKNVETPRSLKLAQHKLNHGGSBAYQCDHGGCFKNPQTWSVLOFH--IKOSH	196
Dp	152	PYKCECGKAFNNSILITGHKIIHTGEKPYKC--GGKAFSKSVTLNTHKAIHAEKRP	208
Qy	197	LKCPKGGKGVGGKKGLSSHMLSHDSTMTKIIYTCYCDVGR-FAKKNELVHNYIIFDGN	255
Dp	209	YKCECGKANSSSSKIMEHRKIHTGEKPYKCEC--GGAFFSSSSTLEHKKI-HAGE	262
Qy	256	IPDDLKETEVEVKLEMLIDQSGKTLNHLHLEITFEKIAVEE-DEADEDSIDLEKSDVSDS	314
Dp	263	KP---YKCECGCAFTWSSSFTTKRIHAAE-KPYKCECGCGFTSILTKHKIIHTGE	318
Qy	315	--MSAQRSITSFPAISLEGSKSVSKLISNGKKLINPRKNCDBMRREYDLRRHLKMH	369
Dp	319	KRYKCECGCAAF--SWSSILITGHKIIHTGEKPYKC--EECGKAFSSRSSSLTRHRIH	371

```

RESULT 5
US-10-029-386-33849
: Sequence 33849, Application US/10029386
: Publication No. US20030194704A1
: GENERAL INFORMATION:
: APPLICANT: Penn, Sharon G.
: APPLICANT: Rank, David R.
: TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR G
: TITLE OF INVENTION: EXPRESSION ANALYSIS TWO
: FILE REFERENCE: AEOMICA-X-2
: CURRENT APPLICATION NUMBER: US/10/029,386
: CURRENT FILING DATE: 2001-12-20
: NUMBER OF SEQ ID NOS: 34288
: SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
: SEQ ID NO 33849
: LENGTH: 487
: TYPE: PRT
: ORGANISM: Homo sapiens
: FEATURE:

```

```

: OTHER INFORMATION: MAP TO AC011467.3
:
: OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 0.67
:
: OTHER INFORMATION: SWISSPROT HIT: Q03923, EVALU0 0.00e+00
US-10-029-386-33849

```

Query Match	17.9%	Score 398	DB 14	Length 487
Best Local Similarity	30.2%	Pred. No. 8.7e-21		
Matches 121; Conservative	48;	Mismatches 167;	Indels 64;	Gaps 18;

```

QY      20  SRPKATICYECCDAVYRPSLLEOHLNTHSNDRPYKCTVDDCKAFRRKSHLTHVSH  79
Db      120  SGKAPKVC--KCGKAYNETHSLTHKRIHNGKKPYKC--BECGKANRSLHTLTKHIIH  175
QY      80  SEKKPFCGVCCKGVNSRHLKRNHEITVT--KSFCTFENCQEAFFYHQSL-RHHILSVH  136
Db      176  TGKKPYKCECGKAFNQSANLTTHKRIHTGKKPYKC--EECGRAFQSSLTTHAKIITHAG  233
QY      137  EKLTLTKQCNKVFTPRSKLAQHKLNHGGSGAYQCDHGGCPKYNQVTNVSUQFH--IKOSH  194
Db      234  EKPYKCECGKAFQSSLTTHKTI-IHTGKKPYKCEB--CGKAFQSSLTTHKRIHSGE  290
QY      195  PKLKCGCKGCGVCKKGGKSSHMLSHSDBTMIKIMTCOYCDVGRKAKNDELVEHYNIFHDG  254
Db      291  KPYKCECGKAFKQSSLTTHKRIHAGE--KFKYCEVCSKA-PSRSHLTTHKRI-HTG  345
QY      255  NIPDDLTKTEVYKLTENLIDQGSKLNMLHELETEKLVKEEDBEDBEDSLDEKSDVYSDS  314
Db      346  EKR---YKCECGKAFNLSQSLTTHKIIHTEB-KPYKCEB-----GKAFNQSST  391
QY      315  MSAQRSI-----KSFNLSEGSXSVSLISNSGKINCPKANNCDRMFSHEYD  361
Db      392  LSHGKVIHTGKRPYKCECGKAFNQS--SHLTTHMIMHTGKKPYKC--EECGKAFNNSSI  447
QY      362  LRHHLKMD-----DLQRIEESFLMSIEKEB  387
Db      448  LNRHKMIHTGKLYKPESCNNACODIATISKYKRNCAEG  487

```

```
RESULT: 6
US-10-734-564-83
: Sequence 83, Application US/10734564
: Publication No.: US20040157278A1
: GENERAL INFORMATION:
: APPLICANT: Christopher C Burgess et al
: TITLE OF INVENTION: Detection Methods Using T1MP
: FILE REFERENCE: 1657//2012
: CURRENT APPLICATION NUMBER: US/10/734,564
: CURRENT FILING DATE: 2003-12-12
: NUMBER OF SEQ ID NOS: 138
: SOFTWARE: FastSeq for Windows Version 4.0
: SEQ ID NO 83
: LENGTH: 423
: TYPE: PRT
: ORGANISM: Homo sapiens
US-10-734-564-83
```

[illegible]

```

Db      247 ----YVC--QKGSFVAKTTELKHAETHKEIILCEVCRKTFKRDYIKQHNKTHAPER 301
      197 --LKCPK--CGKGCYVKGKGLSSHMLS-HDDSTMIKWTCDYCDVGK-FAKKNELVEHYNI 250
      302 DVCRCRGECGGRYTTVFNLOSHILSFHES---RFVCEHAGCGCTFAMKQSLTTH-AV 357
Qy      251 FHDGNIPDDLKETEYVK 268
      358 VHD---PDKKKMKLKVKK 372

RESULT 7
US-10-106-698-5955
; Sequence 5955, Application US/10106698
; Publication No. US20030109690A1
; GENERAL INFORMATION:
; APPLICANT: Ruben et al.
; TITLE OF INVENTION: Colon and Colon Cancer Associated Polynucleotides and Polypeptide
; FILE REFERENCE: PA005PI
; CURRENT APPLICATION NUMBER: US/10/106,698
; PRIOR FILING DATE: 2002-03-27
; PRIOR APPLICATION NUMBER: PCT/US00/26524
; PRIOR FILING DATE: 2000-09-28
; PRIOR APPLICATION NUMBER: US 60/157,137
; PRIOR FILING DATE: 1999-09-29
; PRIOR APPLICATION NUMBER: US 60/163,280
; PRIOR FILING DATE: 1999-11-03
; NUMBER OF SEQ ID NOS: 8564
; SOFTWARE: PatentIn Ver. 3.0
; SEQ ID NO 5955
; LENGTH: 409
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: MISC_FEATURE
; LOCATION: (3)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-10-106-698-5955

Query Match      17.7%; Score 395.5; DB 14; Length 409;
Best Local Similarity 31.1%; Pred. No. 1.1e-20;
Matches 99; Conservative 54; Mismatches 96; Indels 69; Gaps 16;

Qy      8 KAISLS-----ISSSSSR-----PKYICTYEGCDKAYNRPSSLLEOHLRTSNDR 53
      53 ESVSSLITADAFIAGESSAPTPPRPALRRRIFCSPDGSANYSKAMKDAHLCKHTGER 112
Qy      54 PKCTVDDDDKAFPRKSHLETHIVSHSEKKPFGCV--CGKGVNSRQHLKRH-ETH--- 107
      113 PFCVDYEGGKAFIRDYHLSRHILTHTEGKPFVCAMGCDOKFNTKSNLKHFERGHEHQ 172
Qy      108 TSPFKTFENCOEAFYKHOSLR-NHILSVHEKTLTCKO--CNKVFPRPSKLAQHLKXHG 164
      173 QKQYICFSDCKCTFKHQQLKHOCQHTNEPLFKCTQCGCGHGFASPSKLKXHAQAHG 232
Qy      165 GSPAYCOCHPGCFKFOFWSVLQFHIKOSHPRK----- 196
      233 ----YVC--QKGSFVAKTTELKHAETHKEIILCEVCRKTFKRDYIKQHNKTHAPER 287
Qy      197 --LKCPK--CGKGCYVKGKGLSSHMLS-HDDSTMIKWTCDYCDVGK-FAKKNELVEHYNI 250
      288 DVCRCRGECGGRYTTVFNLOSHILSFHES---RFVCEHAGCGCTFAMKQSLTTH-AV 343
Qy      251 FHDGNIPDDLKETEYVK 268
      344 VHD---PDKKKMKLKVKK 358

Db

```

```

; GENERAL INFORMATION:
; APPLICANT: Ghosh, Soumitra S.
; APPLICANT: Fahy, Colin D.
; APPLICANT: Zhang, Bing
; APPLICANT: Gibson, Bradford W.
; APPLICANT: Taylor, Steven W.
; APPLICANT: Glenn, Gary W.
; APPLICANT: Warnock, Dale E.
; TITLE OF INVENTION: TARGETS FOR THERAPEUTIC INTERVENTION
; FILE REFERENCE: 660088.465
; CURRENT APPLICATION NUMBER: US/10/408,765A
; CURRENT FILING DATE: 2003-04-04
; NUMBER OF SEQ ID NOS: 3077
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2105
; LENGTH: 1191
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-408-765A-2105

Query Match      17.7%; Score 395.5; DB 16; Length 1191;
Best Local Similarity 35.8%; Pred. No. 4e-20;
Matches 98; Conservative 40; Mismatches 109; Indels 27; Gaps 12;

Qy      20 SRPKYICTYEGCDKAYNRPSSLLEOHLRTSNDREPKYCTVDDCDKAFPRKSHLETHIVSH 79
      905 TREKTYKC--EECGKAFQPSHLTHHKMHTGEKEYKC--EECGKAFQSSLTTHKILH 960
Qy      80 SEKRPFGSVCGKGVNSRQHLKRHEITHT--KSPFCTENCOEAFYKQSL-RHILSLNH 136
      961 TGEKPKYKCECGKAFRKRSSTLTENKIHTGEKPYKC--EECGKAFQSSLTTHRHRMTG 1018
Qy      137 EKLTCCKCNKVFTRPSKLAQHLKHHGSPAYOCDFGNFOTWGLQFHIKOSHPRK 196
      1019 EKPYKCECGKAFNRSSLTTHKILHTGEKP-YKCEB--CGAFISSSLTNGH-KRIHTR 1074
Qy      197 --LKCPKGCYVKGKGLSSHMLS-HDDSTMIKWTCDYCDVGKFAKKNELVEHYNIFPD 253
      1075 EKPYKCECGKAFQSSLTTHKRLH---TGEKPYKCEB--GKAFKSSALTTHKILHT 1129
Qy      254 GNIPDDLKETEYVKLEMLDOGSTLNALHELET 287
      1130 GE-----KPYKCECGKAFQSSLTTHKILHT 1157
Db

RESULT 9
US-10-203-052B-8
; Sequence 8, Application US/10203052B
; Publication No. US20040109854A1
; GENERAL INFORMATION:
; APPLICANT: Komed Co., Ltd.
; APPLICANT: Lee, Zang Hee
; APPLICANT: Kim, Hong Hee
; TITLE OF INVENTION: A Novel TRAF6 Inhibiting Protein
; FILE REFERENCE: 59520-00002
; CURRENT APPLICATION NUMBER: US/10/203,052B
; CURRENT FILING DATE: 2003-11-03
; NUMBER OF SEQ ID NOS: 52
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 8
; LENGTH: 595
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-203-052B-8

Query Match      17.6%; Score 392.5; DB 16; Length 595;
Best Local Similarity 31.7%; Pred. No. 2.8e-20;
Matches 120; Conservative 48; Mismatches 160; Indels 51; Gaps 17;

Qy      10 ISSLSISS--SRPKYICTYEGCDKAYNRPSSLLEOHLRTSNDREPKYCTVDDCDKAFPR 68
      186 ISCLTEHSRIHTRVNFYKC--EECGKAFNWSSTLTTHKRIHTEKPYKC--EECGKAFNQ 241
Db

```

Qy	69	KSHLETHIVASHSEKPEPHCSVCGKGVNSQH,KRHEITHT--KSPKCFENKCOAPYKHQ	126
Db	242	SSNLIKTKKTHHTEKPEYKCECGKAFNRSTLTHKLIHTEGKPYKC--KECGAFNRSS	299
Qy	127	SL-RHHLIVHEKTLTKOCNKVETPRPSKLAOHKLKHGGSPAYOCDHGPFCKPNQWTSV	169
Db	300	TLTTHRKLIHTEGKPYKCEBGCAFNRSSNLTHKLIHTEGEP-YKCEK--CGKAFNOSAH	356
Qy	186	LQPH--IKOSHPLYKCPKCGKGVCGKGLSSHMLSHDSTMIKIWTCDYCVGKFAKNE	243
Db	357	LTJHEVLIHTEKPEYKCECGKAFNRHSHLTHKLIH--TEBKPYKCEC--GKAFHSS	411
Qy	244	LVEHYNIFHDGPNIPDLLKATEYKXKLEMLLDGSKLANLHLEITPKLVEEDEDEEDSL	303
Db	412	TLTHKLIHNGE-----KPYKCECEKAFNOSKYL-----TEHKLIHTEGEPYE---	455
Qy	304	DEKSDVDRSDSMSAQRISKFTASL-----ESK-----SVBKLISSNGKXINCPKN	355
Db	456	CEKCGKAFNOSSNLTHKRSHTHEKPYKCECGKGFMPSTLTHKLIHTEGKPYKC--E	513
Qy	351	NCDRMFGRBYDLRRHLKWH	369
Db	514	ECGKAFNOSKLTTRKKIHH	532

```

RESULT 10
US-10-094-749-2331
; Sequence 2331, Application US/10094749
; Publication No. US20030219741A1
; GENERAL INFORMATION:
; APPLICANT: ISOGAI, TAKAO
; APPLICANT: SUGIYAMA, TOMOYASU
; APPLICANT: OTSUKI, TETSUJI
; APPLICANT: MAKAMATSU, AI
; APPLICANT: SATO, HIROYUKI
; APPLICANT: ISHII, SHIZUKO
; APPLICANT: YAMAMOTO, JUN-ICHI
; APPLICANT: ISONO, YUUKO
; APPLICANT: HIO, YURI
; APPLICANT: OTSUKA, KAORU
; APPLICANT: NAGAI, KEIICHI
; APPLICANT: IRIE, RYOTARO
; APPLICANT: TAMECHIKA, ICHIRO
; APPLICANT: SEKI, NAOHICO
; APPLICANT: YOSHIKAWA, TSUTOMU
; APPLICANT: OTSUKA, MOTOTYUKI
; APPLICANT: NAGAHARI, KENJI
; APPLICANT: MASUHO, YASUHIKO
; TITLE OF INVENTION: NOVEL FULL-LENGTH CDNA
; FILE REFERENCE: 084335/0160
; CURRENT APPLICATION NUMBER: US/10/094,749
; CURRENT FILING DATE: 2002-03-12
; PRIOR APPLICATION NUMBER: 60/350,435
; PRIOR FILING DATE: 2002-01-24
; PRIOR APPLICATION NUMBER: JP 2001-328381
; PRIOR FILING DATE: 2001-09-14
; NUMBER OF SEQ ID NOS: 3381
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2331
; LENGTH: 555
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-094-749-2331

Query Match 17.6%; Score 392; DB 15; Length 555;
Best Local Similarity 32.5%; Pred. No. 2.8e-20;
Matches 116; Conservative 45; Mismatches 162; Indels 34; Gaps 17,

02 21 RPKKYICITCEGCDKAYNRPSSLLCEHLRTHSNDPRPKYCVDDCDKAFPKSHLEHTIVSHS 80
Db 169 RENSVCQ--EECDCKVFKKPSLTTRKRVHTSEKPKC--EECGKAFKHSSTLTTHKMHT 224

```

[illegible]

```

RESULT 11
US-10-108-260A-3669
; Sequence 3669, Application US/10108260A
; Publication No. US20040005560A1
; GENERAL INFORMATION:
; APPLICANT: HELIX RESEARCH INSTITUTE
; TITLE OF INVENTION: NO. US20040005560A1el full length cDNA
; FILE REFERENCE: H1-A0106
; CURRENT APPLICATION NUMBER: US/10/108,260A
; CURRENT FILING DATE: 2002-03-27
; NUMBER OF SEQ ID NOS: 5458
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3669
; LENGTH: 592
; TYPE: prt
; ORGANISM: Homo sapiens
US-10-108-260A-3669

```

Query Match	17.5%	Score 391	DB 15	Length 592
Best Local Similarity	31.1%	Pred. No. 3,66-20		
Matches	120	Conservative	41	Mismatches 161; Indels 64; Gaps 19;
Qy	23	KKIYITCEGCDKAYNRPSLLEOHLRTSHNDPRYKCTVDDCDKAFPRKSHLETHIVSHSEK	82	
Db	199	KPYIC--EECGAKAFKYSALNTHTKRIHTGEKPYKC--DKCKAFIASSLTSKHEIHTGK	254	
Qy	83	KPHICSVCGKVNRRQHLKXHEIHT--KSKKCTFFENQGEAFYHGQSLRH-----	131	
Db	255	KPYKCECGKAFNOSTLTKHKIKHTEKPYKC--EECGAFNOSTLTKHKIKHTEGKP	312	
Qy	132	-----ILSVH-----EKTLLTCOKCNKFTPRSKLAQHLKHNGSPAYQC	171	
Db	313	YVCECGKAFKYSILTHTKRIHTEKPYKCNKKGKAFIASSLTSRHEFIHM-GKHNYKC	371	
Qy	172	DHPCGFPNQTWSVLQEH--IKOSHPLYKCEPKCGKGVYGGKGLSNHLSHDSMTMICKWT	229	
Db	372	EE--CGKAFIWSVLTTHKRVTHTGEKPYKCECGKAFVASSLTSKHEIIN--TGKKPYK	426	
Qy	230	CDYCDVAG-FAKKNELVEHVIIPDGNIIPDILTETEVKLENLIDGOSKLNMLHELET-	287	
Db	427	CEEC--GKAFQSSSLTKHKKI-HTGEK--YVCECGKAFN--QSSSLTKHKIKHITG	477	
Qy	288	EKLVEDEDEDEDS---LDEKRSDVRSDSMSAORSIKSFTASLEGSKSVSLKINSKG	343	
Db	478	EKPKECECGKAFNOSTLTKHKIKHTEKPYKCECGKAF--HLSTHLTKHKIKHITGK	535	
Qy	344	KINCPKNNCDRMFSREYDLRHLKMH	369	
Db	536	PYRC--RECGKAFNHSATLSHKKIH	559	

RESULT 12

```
US-10-104-047-3748
; Sequence 3748, Application US/10104047
; Publication No. US20030236392A1
; GENERAL INFORMATION:
; APPLICANT: HELIX RESEARCH INSTITUTE
; TITLE OF INVENTION: No. US20030236392A1e1 full length cDNA
; FILE REFERENCE: H1-A0105
; CURRENT APPLICATION NUMBER: US/10/104,047
; PRIOR FILING DATE: 2002-03-25
; PRIOR APPLICATION NUMBER:
; NUMBER OF SEQ ID NOS: 4096
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3748
; LENGTH: 540
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-104-047-3748

Query Match      17.5%; Score 390; DB 15; Length 540;
Best Local Similarity 30.0%; Pred. No. 3,8e-20;
Matches 107; Conservative 61; Mismatches 155; Indels 34; Gaps 15;

QY 23 KYVICGECDCAYNRPSLLEQHLRTHSNDRPYKCTVDDCDKAFPRKSHLETHIVSHSRK 82
DB 198 KLYKC--QECDFTFNQFSNLTVEKDYAREKPYKC--EECGKAFNQSHTLTHKIIHTGE 253
QY 83 KEFHSCVCGGVNSRQHLKRHEITHTKSFCTFENCQBAFYKHQSL-RHHILSVHEKTLT 141
DB 254 KKYKECECGKAFNQSHTLTHKIIHTGEBOPIYCEEGKAFNQSHTLTHKIIHTGEKPYK 313
QY 142 CKQCNKVFTRPSKLAQHLKHHGSPAYOCDHGCGKFNQFWSVLQFH--IKOSHFKLKC 199
DB 314 CECEGKAFNRSSKLTGKHIHTGECP--YKCEB--CGKAFNRSSNLTGKHIHTGEKPYKC 370
QY 200 PRCGCGVCGKGLSHMLSHDSTMTKIWTCDYCDVGR-FAKNLEVEHYNIHFHGNID 258
DB 371 KECGKAFNRSSALTTHKRIH--TGKPYKCEB--GKAFNRSSNLTGKHIHTGEKPYKC 421
QY 259 DLLKTEVYKLENNLDQSKLNNLHLEFTEKLVKVEDE-----EDEEDSLDEKSPDRSD 313
DB 422 ---KPYKECECGKAFNQSHTLTHKIIHTGEBOPIYKEEGKAFKHSSTLTHKRIHTGEK 478
QY 314 SMSAQRISFTASLEGSKVSLSNSGKK-INCPKNNCDMPFSREYDLRRLKWH 369
DB 479 PYKCECGKAFNRS---SKLTGKHIHTGEKPYKCBR--CDKAFNQSANLTGKRIH 530

RESULT 13
US-10-094-749-2365
; Sequence 2365, Application US/10094749
; Publication No. US20030219741A1
; GENERAL INFORMATION:
; APPLICANT: ISOGAI, TAKAO
; APPLICANT: SUGIYAMA, TOMOYASU
; APPLICANT: OTSUKI, TETSUJI
; APPLICANT: WAKAMATSU, AI
; APPLICANT: SATO, HIROYUKI
; APPLICANT: ISHII, SHIZUKO
; APPLICANT: YAMAMOTO, JUN-ICHI
; APPLICANT: ISONO, YUUKO
; APPLICANT: HIO, YURI
; APPLICANT: OTSUKA, KAORU
; APPLICANT: NAGAI, KEIICHI
; APPLICANT: IRIE, RYOTARO
; APPLICANT: TAMECHIKA, ICHIRO
; APPLICANT: SEKI, NAOHITO
; APPLICANT: YOSHIKAWA, TSUTOMU
; APPLICANT: OTSUKA, MOTOUYUKI
; APPLICANT: NAGAHARA, KENJI
; APPLICANT: MASUHO, YASUHIKO
; TITLE OF INVENTION: NOVEL FULL-LENGTH CDNA
; FILE REFERENCE: 08435/0160
```

```
US-10-104-047-3748
; Sequence 3748, Application US/10104047
; Publication No. US20030236392A1
; GENERAL INFORMATION:
; APPLICANT: HELIX RESEARCH INSTITUTE
; TITLE OF INVENTION: No. US20030236392A1e1 full length cDNA
; FILE REFERENCE: H1-A0105
; CURRENT APPLICATION NUMBER: US/10/104,047
; PRIOR FILING DATE: 2002-03-25
; PRIOR APPLICATION NUMBER:
; NUMBER OF SEQ ID NOS: 4096
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3748
; LENGTH: 540
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-104-047-3748

Query Match      17.5%; Score 390; DB 15; Length 540;
Best Local Similarity 30.0%; Pred. No. 3,8e-20;
Matches 107; Conservative 61; Mismatches 155; Indels 34; Gaps 15;

QY 23 KYVICGECDCAYNRPSLLEQHLRTHSNDRPYKCTVDDCDKAFPRKSHLETHIVSHSRK 82
DB 198 KLYKC--QECDFTFNQFSNLTVEKDYAREKPYKC--EECGKAFNQSHTLTHKIIHTGE 253
QY 83 KEFHSCVCGGVNSRQHLKRHEITHTKSFCTFENCQBAFYKHQSL-RHHILSVHEKTLT 141
DB 254 KKYKECECGKAFNQSHTLTHKIIHTGEBOPIYCEEGKAFNQSHTLTHKIIHTGEKPYK 313
QY 142 CKQCNKVFTRPSKLAQHLKHHGSPAYOCDHGCGKFNQFWSVLQFH--IKOSHFKLKC 199
DB 314 CECEGKAFNRSSKLTGKHIHTGECP--YKCEB--CGKAFNRSSNLTGKHIHTGEKPYKC 370
QY 200 PRCGCGVCGKGLSHMLSHDSTMTKIWTCDYCDVGR-FAKNLEVEHYNIHFHGNID 258
DB 371 KECGKAFNRSSALTTHKRIH--TGKPYKCEB--GKAFNRSSNLTGKHIHTGEKPYKC 421
QY 259 DLLKTEVYKLENNLDQSKLNNLHLEFTEKLVKVEDE-----EDEEDSLDEKSPDRSD 313
DB 422 ---KPYKECECGKAFNQSHTLTHKIIHTGEBOPIYKEEGKAFKHSSTLTHKRIHTGEK 478
QY 314 SMSAQRISFTASLEGSKVSLSNSGKK-INCPKNNCDMPFSREYDLRRLKWH 369
DB 479 PYKCECGKAFNRS---SKLTGKHIHTGEKPYKCBR--CDKAFNQSANLTGKRIH 530

RESULT 13
US-10-094-749-2365
; Sequence 2365, Application US/10094749
; Publication No. US20030219741A1
; GENERAL INFORMATION:
; APPLICANT: ISOGAI, TAKAO
; APPLICANT: SUGIYAMA, TOMOYASU
; APPLICANT: OTSUKI, TETSUJI
; APPLICANT: WAKAMATSU, AI
; APPLICANT: SATO, HIROYUKI
; APPLICANT: ISHII, SHIZUKO
; APPLICANT: YAMAMOTO, JUN-ICHI
; APPLICANT: ISONO, YUUKO
; APPLICANT: HIO, YURI
; APPLICANT: OTSUKA, KAORU
; APPLICANT: NAGAI, KEIICHI
; APPLICANT: IRIE, RYOTARO
; APPLICANT: TAMECHIKA, ICHIRO
; APPLICANT: SEKI, NAOHITO
; APPLICANT: YOSHIKAWA, TSUTOMU
; APPLICANT: OTSUKA, MOTOUYUKI
; APPLICANT: NAGAHARA, KENJI
; APPLICANT: MASUHO, YASUHIKO
; TITLE OF INVENTION: NOVEL FULL-LENGTH CDNA
; FILE REFERENCE: 08435/0160

US-10-203-052B-7
; Sequence 7, Application US/10203052B
; Publication No. US20040109854A1
; GENERAL INFORMATION:
; APPLICANT: KOMED CO., LTD.
; APPLICANT: Lee, Zang Hee
; APPLICANT: Kim, Hong Hee
; TITLE OF INVENTION: A Novel TRAF6 Inhibiting Protein
; FILE REFERENCE: 59520-00002
; CURRENT APPLICATION NUMBER: US/10/203,052B
; PRIOR FILING DATE: 2003-11-03
; NUMBER OF SEQ ID NOS: 52
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 7
; LENGTH: 568
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-203-052B-7

Query Match      17.4%; Score 388; DB 16; Length 568;
Best Local Similarity 28.5%; Pred. No. 5,7e-20;
Matches 105; Conservative 62; Mismatches 167; Indels 34; Gaps 13;

QY 15 SSSSSRPPKTYIC---TYEGCDKAYNRPSLLEQHLRTHSNDRPYKCTVDDCDKAFPRKS 70
DB 119 BCDKAFNQSHTLTHKIIHTGEKPYKCECGKAFNQSHTLTHKIIHTGEKPYKC--EECGK 274
QY 63 DKAFPRKSHLETHIVSHSRKPFHCSVCGGVNSRQHLKRHEITHT--KSFCTFENCQOE 120
DB 217 BCDKAFNQSHTLTHKIIHTGEKPYKCECGKAFNQSHTLTHKIIHTGEKPYKC--EECGK 274
QY 121 AFYKHQSL-RHHILSVHEKTLTKQCNKVFTRPSKLAQHLKHHGSPAYOCDHGCGKFN 179
DB 333 AFNKSHTLTHKRIHTGEKPYKCECGKAFNQSHTLTHKIIHTGEKPYKC--CGKA 389
QY 180 FQTSVLQFH--IKOSHFKLKCPCGKCGVCGKGLSHMLSHDSTMTKIWTCDYCDVGR 237
DB 390 FQFQSNLTTHKRIHTGEKPYKCECGKAFNQSHTLTHKRIH--TGKSKYKCEB--GK 444
QY 238 -FAKNLEVEHYNIHFHGNIDDLLKTEVYKLENNLDQSKLNNLHLEFTEKLVKVEDE 296
DB 445 AFYRSSLTHKRIHTGEKPYKCECGKAFNQSHTLTHKIIHTGEKPYKC--CGKA 490
QY 297 EDEEDSLDEKSPDRSDSMSAQRISFTASLEGSKVSLSNSGKKINCPKNNCDMPF 356
DB 491 KPYQ-----CECGKAFNQSHTLTHKRIHTGEKPYKCECGKAFNQSHTLTHKRIHTGEK 530
QY 357 SREYDLRRLKWHNDLQRISSFLNSIKEETPRGEPYVKARNDLLENETSVISR 412
DB 531 NQSSNLTGKRIH-----TGKLYKPRKCNDSDENTSKFSK 566

RESULT 14
US-10-203-052B-7
; Sequence 7, Application US/10203052B
; Publication No. US20040109854A1
; GENERAL INFORMATION:
; APPLICANT: KOMED CO., LTD.
; APPLICANT: Lee, Zang Hee
; APPLICANT: Kim, Hong Hee
; TITLE OF INVENTION: A Novel TRAF6 Inhibiting Protein
; FILE REFERENCE: 59520-00002
; CURRENT APPLICATION NUMBER: US/10/203,052B
; PRIOR FILING DATE: 2003-11-03
; NUMBER OF SEQ ID NOS: 52
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 7
; LENGTH: 568
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-203-052B-7

Query Match      17.4%; Score 388; DB 16; Length 568;
Best Local Similarity 28.5%; Pred. No. 5,7e-20;
Matches 105; Conservative 62; Mismatches 167; Indels 34; Gaps 13;

QY 15 SSSSSRPPKTYIC---TYEGCDKAYNRPSLLEQHLRTHSNDRPYKCTVDDCDKAFPRKS 70
DB 119 BCDKAFNQSHTLTHKIIHTGEKPYKCECGKAFNQSHTLTHKIIHTGEKPYKC--EECGK 274
QY 63 DKAFPRKSHLETHIVSHSRKPFHCSVCGGVNSRQHLKRHEITHT--KSFCTFENCQOE 120
DB 217 BCDKAFNQSHTLTHKIIHTGEKPYKCECGKAFNQSHTLTHKIIHTGEKPYKC--EECGK 274
QY 121 AFYKHQSL-RHHILSVHEKTLTKQCNKVFTRPSKLAQHLKHHGSPAYOCDHGCGKFN 179
DB 333 AFNKSHTLTHKRIHTGEKPYKCECGKAFNQSHTLTHKIIHTGEKPYKC--CGKA 389
QY 180 FQTSVLQFH--IKOSHFKLKCPCGKCGVCGKGLSHMLSHDSTMTKIWTCDYCDVGR 237
DB 390 FQFQSNLTTHKRIHTGEKPYKCECGKAFNQSHTLTHKRIH--TGKSKYKCEB--GK 444
QY 238 -FAKNLEVEHYNIHFHGNIDDLLKTEVYKLENNLDQSKLNNLHLEFTEKLVKVEDE 296
DB 445 AFYRSSLTHKRIHTGEKPYKCECGKAFNQSHTLTHKIIHTGEKPYKC--CGKA 490
QY 297 EDEEDSLDEKSPDRSDSMSAQRISFTASLEGSKVSLSNSGKKINCPKNNCDMPF 356
DB 491 KPYQ-----CECGKAFNQSHTLTHKRIHTGEKPYKCECGKAFNQSHTLTHKRIHTGEK 530
QY 357 SREYDLRRLKWHNDLQRISSFLNSIKEETPRGEPYVKARNDLLENETSVISR 412
DB 531 NQSSNLTGKRIH-----TGKLYKPRKCNDSDENTSKFSK 566
```

Db 212 SSKLTHKRIYTCCKLYKCOECDRTFNQSNLTYYKKDYAREKPYK--EECGKAFNQSS 269
QY 71 HLETHIVSHSEKPPHCOSVCGKVNRSQHLKREIHTTSCFKCTFENCQAFYKHQSL-R 129
Db 270 HLTHKLIHTGKPYKCEBGKASNOFSNLTTHKLIHTGQOPYICTECCGAFQSSSTLT 329
QY 130 HHILSVHEKTLTCKQCNKVFTRPSPKLAQHLKHHGSPAYQCDHPGCFKXFTWVSLQFH 189
Db 330 HKRIHTGKPYKCEBGKANNRSSKLTENKNIHTGEPR-YKCEB--CGKAFNNSULTEH 366
QY 190 --IKOSHPRKLCPCCKGCVGKKGLSSHMLSHDOSTMIKTIWTCDDYCVGFAKKNELVEH 247
Db 387 RKIHTEEKPYKCEBGKAFNNSALTTTHKRIH--TGEKPYKCEB--CGKAFNNSULTE 441
QY 248 YNIFHDGNIPTDLLKETEYVKLENLDDQSKLNNHLETEKLVBEDE----EDBEDS 302
Db 442 YKLIHTGK-----KPYKCEBGKAFIQSSKLTENKNIHTGELPYKCEBGKAFKSSSL 495
QY 303 LDEKSDVRSDSMSAORSIKSFTASLEGSKVSGLISNSGKK--INCPKNNCDRMPFREVYD 361
Db 496 THKRIHTGKPYKCEBGKAFNNS--SKLTENKNIHTGKPYKCEB--CDKAFNNSAN 550
QY 362 LRRHLKWH 369
Db 551 LTKGKKIH 558

RESULT 15
US-10-029-386-31985
/ Sequence 31985, Application US/10029386
/ Publication No. US20030194704A1
/ GENERAL INFORMATION:
/ APPLICANT: Penn, Sharon G.
/ APPLICANT: Rank, David R.
/ APPLICANT: Hanzel, David K.
/ TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR G
/ TITLE OF INVENTION: EXPRESSION ANALYSIS TWO
/ FILE REFERENCE: AEOMICA-X-2
/ CURRENT APPLICATION NUMBER: US/10/029,386
/ CURRENT FILING DATE: 2001-12-20
/ NUMBER OF SEQ ID NOS: 34288
/ SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
/ SEQ ID NO 31985
/ LENGTH: 742
/ TYPE: PRT
/ ORGANISM: Homo sapiens
/ FEATURE:
/ OTHER INFORMATION: MAP TO AC010328.4
/ OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 1.7
/ OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 1.8
/ OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 1.8
/ OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 1.5
/ OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 1.7
/ OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 1.5
/ OTHER INFORMATION: SWISSPROT HIT: Q05481, EVALU 0.00e+00
US-10-029-386-31985

Query Match 17.4%; Score 387.5; DB 14; Length 742;
Best Local Similarity 30.7%; Pred. No. 8.6e-20;
Matches 112; Conservative 46; Mismatches 154; Indels 53; Gaps 16;

QY 11 SSI-SSSSSRPKYICTYEGCDKAVNRPSLEQHLKRTSNDRPYKCTVDDCDKAFPRK 69
Db 289 SSIAHQATHSGKPYK--NECGKVFQNSHLTNHRIHTGKPYK--NECGKAFGVR 344
QY 70 SHLETHIVSHSEKPPHCOSVCGKVNRSQHLKREIHTT--KSFCTFENCQAFYKHQSL 127
Db 345 SSIAHQATHSGKPYKCEBGKANNRSSKLTENKNIHTGEPR-YKCEB--CGKAFNNSULTEH 402
QY 128 L-BHHILSVHEKTLTCKQCNKVFTRPSPKLAQHLKHHGSPAYQCDHPGCFKXFTWVSL 186
Db 403 LTHQVYIHTGKPYKCEBGKAFNNSALTTTHKRIH--TGEKPYKCEB--CGKAFNNSUL 459

QY 187 OFH--IKOSHPRKLCPCCKGCVGKKGLSSHMLSHDOSTMIKTIWTCDDYCVGFAKKNEL 244
Db 460 THQVYIHTGKPYKCEBGKANNRSSKLTENKNIHTGEPR-YKCEB--CGKAFNNSULTEH 514
QY 245 YNIFHDGNIPTDLLKETEYVKLENLDDQSKLNNHLETEKLVBEDEDEBEDSLD 304
Db 515 LSRHQRIHTGKPYKCEBGKANNRSSKLTENKNIHTGEPR-YKCEB--CGKAFNNSULTEH 560
QY 305 EKRSVDVRSDSMSAORSIKSFTASLEGSKVSGLISNSGKK--INCPKNNCDRMPFREVYD 364
Db 561 -----CGKVFQNSHLAR--HRRVHTGKPYQ--NECGKAFQTSKLAR 601
QY 365 HLKWH 369
Db 602 HQRVH 606

Search completed: February 8, 2005, 15:34:42
Job time : 131 secs

This Page Blank (uspto)


```

Db      141  CNLRPYGHPQLRAHLNLSVHLKLTCTPHCNKSPFQRYRLRNHISKHNDEVENPVOCTAG  200
Qy      176  CFKQFQWVSVLQFHFKQSHPKLCKPCGCKGCVGKGLSSHMLSHDSDTMIKWTCDYCV  235
Db      201  CCKEFRISOLQSHIKNDHPKLCPICSKPCGVEGNGLMHMIHDDSLVTKWKKHICPD  260
Qy      236  GKPAKQNLVHNYNFFH--DGNIPDDL-LKETPEVKLEMLDQGSGLNLLHLETKLVE  293
Db      261  MFSFKSHLLTYGSIHTBEDPLETKYKISDIOQL--VOHGVGLN-----SKHSNE  312
Qy      294  EDEBEDSDLEDEKSDVRSDSMAQRSIKSFASLE--GSKSVSKLISGKINCPKNC  352
Db      313  QDEEKISRLKRRKRLTENNNVPEFLQNEVDLEKRESENGNLNLLNTVGRKRYRFPYNNC  372
Qy      353  DRMSRREYDLRRHL---KMHDDNLQRIESFLNISKEE  387
Db      373  SRTFETKEKYKHIDKHVYHEKLIQ-----EKEE  404

```

RESULT 2
T37676
zinc-finger protein - fission yeast (*Schizosaccharomyces pombe*)
C|Species: Schizosaccharomyces pombe
C|Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 09-Jul-2004
C|Accession: T37676
R|McDonagall, R.C.; Rajandream, M.A.; Barrell, B.G.; Davis, P.; Churcher, C.M.
submitted to the EMBL Data Library, October 1999
A|Reference number: 221736
A|Accession: T37676
A|Status: preliminary; translated from GB/EMBL/DBJ
A|Molecule type: DNA
A|Residues: 1-374 <MCD>
A|Cross-references: UNIPROT:O9UTL5; EMBL:AL132675; PIDN:CA859689.1; GSPDB:GN00066; SPDB
A|Experimental source: Strain 972h-, cosmid c144
C|Genetics:
A|Gene: SPDB:SPAC144..09c
A|Map position: 1
A|Introns: 50/3
C|Superfamily: zinc finger protein ZFP-36; LIM metal-binding repeat homology

Query	March	22.2*	Score 495;	DB 2;	Length 374;
Best Local	Similarity	32.8*	Pred. No. 6.7e-23;		
Matches	123;	Conservative	58;	Mismatches 160;	Indels 34; Gaps 13;
QY	15	SSSSSSRRKKYICYEGGCDKAYNRPSSLLEQHLRTHSNRPYKCTVDDCDKAFFRKSHLET	74		
DB	13	SKNIRSAKKIHCHPEEGCGKKYSRPSLLEQLRHTSHSNRPVCDVYGSKAFYRKSHIKI	72		
QY	75	HIVHSEKKPPHCSV--CGKGVNSRQHLKRHEITH--TKSFKCTFENCQAFYKHQSIIRH	130		
DB	73	HKRCHTNVKKPSCSHYDGCDAQFYTQQLHERHIEVVRKKRPACTWEGCDEFFSKHQQLRS	132		
QY	131	HLISVHEKTL-----TCKQCKNVFTRPSKLAQHLKHHGSGPAYOCDHGGC--FKNPQWMS	184		
DB	133	HISACHTLLPLPYCTYQDCBLRFATKQKLQNVNVRNAREKIIISYSCPHESCVGHEGFEKMS	192		
QY	185	VLAGPHIKOSHEPLKCPKCGKGVGCKGLSSHMLSHDSTMIKWTCDYCDV----GKFAK	240		
DB	193	QLONNHREAPVP--SCSICGRQFKTAALHRRHVLAH--QTLLEERTKYCHPNEGCKKSPTR	249		
QY	241	KNELVEHNINIHDDNIIP---DDLKETEYK--KLENLDDQSKLNNHLELEFEKLKVEBDE	296		
DB	250	SSALKRHLISVHEGMAFHCDSCGTFKGYKMLRPHLRGTF--CKKAHPYINEGCIKIDG	308		
QY	297	EDEEDSLDEKRSVDYSDSMS--AORSIKSFTASLEGSKSVKLSINSGKILNCPNNCDR	354		
DB	309	IEGVALIHQCKKEKELSSNLVSDVAKKIINEVTG--HGKY-----EAREYCSGSPPECNY	358		
QY	355	MFSEREYDLRRHLKWH	369		
DB	359	RFKRLYDMGRHLNSH	373		

RESULT 3
S26823
zinc finger protein ZNF43 - human
N:Alternate names: zinc finger protein kox27
C:Species: Homo sapiens (man)
C:Date: 13-Jan-1995 #sequence revision 13-Jan-1995 #text_change 09-Jul-2004
C:Accession: S26823; I37967; S10416
R:Lovering, R.; Trowdale, J.
Nucleic Acids Res. 19, 2921-2928, 1991
A>Title: A gene encoding 22 highly related zinc fingers is expressed in lymphoid cell li
A:Reference number: S26823; MUID:91279444; PMID:1711675
A:Accession: S26823
A:Status: Preliminary
A:Molecule type: mRNA
A:Residues: 1-803 <LOV>
A:Cross-references: UNIPROT:P17038; EMBL:X59244; NID:G38031; PID:CAA41932.1; PID:G380323
R:Thiesen, H.J.
New Biol. 2, 363-374, 1990
A>Title: Multiple genes encoding zinc finger domains are expressed in human T cells.
A:Reference number: I37949; MUID:91145339; PMID:2288909
A:Accession: I37967
A:Status: Preliminary; translated from GB/EMBL/DBJ
A:Residues type: mRNA
A:Residues: 476-531 <THI>
A:Cross-references: EMBL:X52358; NID:G34160; PID:CAA36584.1; PID:G930090
C:Genetics:
A:Gene: GDB:ZNF43; HMP6
A:Cross-references: GDB:128653
A:Map position: 19p13.1-19p12
C:Superfamily: zinc finger protein ZFP-36; LIM metal-binding repeat homology
C:Keywords: DNA binding; zinc finger

[illegible]

RESULT 4
G02075
transcription repressor zinc finger protein 85 - human
C|Species: Homo sapiens (man)
C|Date: 21-Dec-1996 #sequence_06-Jun-1997 #text_change 09-Jul-2004
C|Accession: G02075
R|Poncellet, D.A.
submitted to the EMBL Data Library, September 1995

```

A:Reference number: G09169
A:Accession: G02075
A:Status: preliminary; translated from GB/EMBL/DBSJ
A:Accession: G02075
A:Molecule type: mRNA
A:Residues: 1-595 <PDB>
A:Cross-references: UNIPROT:Q03923; EMBL:U35376; NID:g1017721; PIDN:AAA79179.1; PID:g101
A:Map position: 19p12-19p12
C:Superfamily: zinc finger protein ZFP-36, LIM metal-binding repeat homology

Query Match
Best Local Similarity 31.7%; Score 396.5; DB 2; Length 595;
Matches 120; Conservative 48; Mismatches 160; Indels 51; Gaps 17;

QY 10 ISSLISSSS-SRPKKYICTYEGCDYAYNRPSSLLEOHLRTHSNDPRYKCTVDDCDKAFPR 68
DB 186 ISCLTSHSRHITHTVNFYK--EECGKAFNWSSTLTGKRIHTGEKPYK--EECGKAFNQ 241

QY 69 KSHLETHIYSHSEKKEFHSCVCGGVNSRQHLKRHEITHT--KSFCTFENCOEAFYKQ 126
DB 242 SSNLINHKIKHTEKPYKCEBCKAFNRPSLTTHKIHTGEKPYK--KECGKAFNRSS 299

QY 127 SL-RHHILSVHEKTLTCKQCNKYFTRPBKLQHLKHHGSPAYQCDHPCCFPNFTWGV 185
DB 300 TLTHHKIHTGEKPYKCEBCKAFKQSSNLTTHKIHTGEKPYK--CGKAFNQSAAH 356

QY 186 LQFH--IKOSHPLKCPKCGKGVGKGLSSHMLSHDDSTMIKIWTCDYCDVGFPAKKE 243
DB 357 LTHVEVHTGEKPYKCEBCKAFNHSHTLTHKIHT--TGEKPYKCKG--GKAFKHS 411

QY 244 LVENHVIHFDGNIPDOLLKETEYKLENDLQSKLNNHLETEKLYAEDEDEDESL 303
DB 412 TLTKHKIHTGE--KPYKKECEKAFNQS--TEHKIHTGEKPYE-- 455

QY 304 DEKRSIVRSDSMSAQSISFTASL-----EGSK-----SVSKLISNGKICPNK 350
DB 456 CEKCGKAFNQSSTLTTHKSHTEKPYKCEBCKGKFRBETLTTHKIHTGEKPYK--E 513

QY 351 NCDRMFSREYDLRRHLKWH 369
DB 514 ECGKAFNQSSTLTTHKSH 532

RESULT 5
138937
DNA/RNA-binding protein - human (fragment)
C:Species: Homo sapiens (man)
C>Date: 29-May-1998 #sequence_revision 29-May-1998 #text_change 21-Jul-2000
C:Accession: I38937
R:Drew, P.D.; Nagle, J.W.; Canning, R.D.; Ozato, K.; Biddison, W.B.; Becker, K.G.
Gene 159, 215-218, 1995
A:Title: Cloning and expression analysis of a human cDNA homologous to Xenopus TFIIIA.
A:Reference number: I38937; MID:95347600; PMID:7622052
A:Accession: I38937
A:Status: preliminary; translated from GB/EMBL/DBSJ
A:Molecule type: mRNA
A:Residues: 1-363 <RES>
A:Cross-references: EMBL:U20272; NID:9644870; PIDN:AAA75623.1; PID:9644871
C:Superfamily: transcription factor IIA

Query Match
Best Local Similarity 17.7%; Score 395.5; DB 2; Length 363;
Matches 99; Conservative 54; Mismatches 96; Indels 69; Gaps 16;

QY 8 KSISSTL-----ISSSSSSR-----PKYICTYEGCDKAYNRPSSLLEOHLRTSNR 53
DB 7 ESIVSLTIDAFIAAGESSAPTPRRPALPRFICSPDCSANYSKAMKIDAHLCIKHTGER 66

QY 54 PYKCTVDDCDKAFPRKSHLETHIYSHSEKKEFHSCV--CGKGVNSRQHLKRH-EITH-- 107
DB 67 PFCVDEYGGCGKAFIRDYHLRHLITHTGEKPYKFCVCAANGCDQKENTSNLKKHFERGHEN 126

```

```

QY 108 TKSFKTFENCOEAFYKQSLR-NHILSVHEKTLTCKQ--CNKVFTRPBKLQHLKHHG 164
DB 127 QKQYICSPFEDCKKTFKQHLKIHQOQHINERPLFKCTQSGCGHFSPSKLRHAAHNE 186

QY 165 GSPAYQCDHPGCFKQFQTVSLQFHKQSHPK----- 196
DB 187 ----YVC-QKGSFVAKTWTTELLKHVETHEILCEVCRKTFPKRDYLLQKMHKTAPE 241

QY 197 --LKCPK--CGKGVCKKSLSSHMLS-HDSTMIKIWTCDYCDVGR-PAKKNELVHNVI 250
DB 242 DVCRCRCEGCGRTYTTVFNLSHILSFHES--RPVCEHAGCGKTFAMKQSLTRH-AV 297

QY 251 FHDGNIPDOLLKETEYK 268
DB 298 VHD--PDKKMKLKYK 312

RESULT 6
535305
zinc finger protein ZNF91 - human
C:Species: Homo sapiens (man)
C>Date: 31-Dec-1993 #sequence_revision 02-Jun-1994 #text_change 09-Jul-2004
C:Accession: S35305
R:Bellefroid, B.J.; Martine, J.C.; Ried, T.; Lecocq, P.J.; Riviere, M.; Amemiya, C.; Pon
EMBO J. 12, 1363-1374, 1993
A:Title: Cloned organization of homologous KRAA zinc-finger genes with enhanced expr
A:Reference number: S35305; MID:93223677; PMID:8467795
A:Molecule type: mRNA
A:Residues: 1-1191 <BEL>
A:Cross-references: UNIPROT:Q05481; EMBL:L11672; NID:g186773; PIDN:AAA59469.1; PID:g186
A:Note: the authors translated the codon GCA for residue 750 as Thr and GCT for residue
C:Keywords: DNA binding; zinc finger

Query Match
Best Local Similarity 17.7%; Score 393.5; DB 2; Length 1191;
Matches 98; Conservative 40; Mismatches 109; Indels 27; Gaps 12;

QY 20 SRPKYICTYEGCDKAYNRPSSLLEOHLRTHSNDPRYKCTVDDCDKAFPRKSHLETHIYSH 79
DB 905 TREKPYK--EECGKAFSQPSHLLTHKRMHTGEKPYK--EECGKAFSQSSTLTTHKIH 960

QY 80 SEKKPFHSCVCGGVNSRQHLKRHEITHT--KSFCTFENCOEAFYKQSL-RHHILSYH 136
DB 961 TGEKPYKCEBCKAFKSSTLTEHKIHTGEKPYK--EECGKAFSQSSTLTTHKIHMTNG 1018

QY 137 EKTTLCKQCNKYFTRPBKLQHLKHHGSPAYQCDHPCCFPNFTWVSLQFHKQSHPK 196
DB 1019 EKPYKCEBCKAFNRSSKLTTHKIHTGEKPYK--YKCEB--CGKAFISSSTLNGH-KRIHR 1074

QY 197 ---LKCPKCGKGVGKGLSSHMLSHDDSTMIKIWTCDYCDVGFPAKKNELVHNVI 253
DB 1075 EKPYKCEBCKAFSQSSTLTTHKRLH--TGEKPYKCGC--GKAFESSALTTHKIH 1129

QY 254 GNIPDOLLKETEYKLENDLQSKLNNHLEET 287
DB 1130 GE-----KPYKCEKCKAFNQSSTLTTHKIH 1157

RESULT 7
500647
finger protein - African clawed frog
C:Species: Xenopus laevis (African clawed frog)
C>Date: 30-Sep-1989 #sequence_revision 30-Sep-1989 #text_change 31-Dec-1993
C:Accession: S00647
R:Ruiz i Altaba, A.; Perry-O'Keefe, H.; Melton, D.A.
EMBO J. 6, 3065-3070, 1987
A:Title: Xfin: an embryonic gene encoding a multifingered protein in Xenopus.

```

A:Reference number: S00647; MUID:88082679; PMID:282129
A:Accession: S00647
A:Molecule type: mRNA
A:Residues: 1-1350 <RUI>
A:Cross-references: EMBL:X06021
A:Note: It is uncertain whether Met-1, Met-11, Met-16, Met-38, or Met-39 is the initiator
C:Genetics:
A:Gene: fin
C:Keywords: DNA binding; zinc finger

Query Match 16.8%; Score 375.5; DB 2; Length 1350;
Best Local Similarity 27.3%; Pred. No. 4.5e-15;
Matches 119; Conservative 65; Mismatches 171; Indels 81; Gaps 20;

QY 11 SSLSSSSSSSPKKKTCYBEGCDKAYNRPSSLEQHLRHSNDPRYKCTVDDCAFFRKS 70
DB 340 SPLNHQCHSREKPYLCSH--CNKGFIQNSDLVHGFRHTGTERPYCA--ECHGFIQKS 395
QY 71 HLEHTIVHSKKPKPHGCVGKGVNSRQHLKRHTHT--RSFKTFCENCOEAFYKQSL 128
DB 396 DLVHKLRTHTGKRPKCSCHCDKFTERSALMHQRTHTGKPYKCS--DCKEFTQNSL 453
QY 129 -RHHILSVAEKTLTCQKCNKVFTRPSKLAQKLYH-----HGSPAYCDDHGFCKN 179
DB 454 ILHQRIHNGEPRYKCTLDRTFIQNSDLVHGKQYHANNPLSDPHTPANSPHKCSK--CDLT 511
QY 180 FQWMSVLQFNIK--QSHDKLKCPKCGKGVCKGLSSHMLSH-----DD 221
DB 512 FSHNSTFPMKSHKLSHGEEKKFOCAECKKGFTQKSLDVKHIRVHTGEEKPKCLLCKKSPSQN 571
QY 222 STMKIKV-----TQDYCDVGKFAKAGELVEHYNI FPHQGNIPDLLKTEVKYKLENL 272
DB 572 SDLHGMHHTGEEKRPFCYTD-KSFTERSALEKHNRT-HTGEBPH-----KCSVCKG 623
QY 273 LDQSKLTNNLHLETEKLVKVEDEDEDEDSLDEKRSQVPSMSAQR-----S 320
DB 624 FIQKSAL-----TKSHRTHTGKRPYCTCGKSPIONSLVHGQRHTGEEKPYHCTEC 676
QY 321 IKSPTASLEGSKSVKSLISNSGKK-INCPKNNCDRMFGREYDLRRHLKMD-DNLQRIES 378
DB 677 NKRFY---EGSSLVVHRRHTSGEEKPYRCFQ--CEKTFIQSSDLVHGLVHNGENPPAPATA 731
QY 379 FLN-SIEKEETPEGEF 393
DB 732 FHEILIRRENTLRSEP 747

RESULT 8
148668
zinc finger protein 51 - mouse
N:Alternate names: finger protein zfc12
C:Species: Mus musculus (house mouse)
C:Date: 02-Jul-1996 #sequence_revision 02-Jul-1996 #text_change 09-Jul-2004
C:Accession: 148668; A40984
R:Burke, P.S.; Don, J.; Wolgemuth, D.J.
Mamm. Genome 5, 387-389, 1994
A:Title: Zfp-51, a murine zinc finger encoding gene mapping to the t-complex region of a
A:Reference number: 148668; MUID:94319090; PMID:8043957
A:Accession: 148668
A:Status: preliminary
A:Status: preliminary
A:Residues: 1-710 <RES>
A:Cross-references: UNIPROT:O61898; EMBL:X74855; NID:9488832; PIDN:CA52847.1; PID:94888
Proc. Natl. Acad. Sci. U.S.A. 88, 7923-7927, 1991
A:Title: A cluster of related zinc finger protein genes is deleted in the mouse embryonic
A:Accession: A40984; MUID:91376058; PMID:1680234
A:Molecule type: mRNA
A:Residues: 672-710 <CRO>
A:Cross-references: GB:M74235
C:Genetics:

A:Gene: zfp-51
 C:Superfamily: zinc finger protein zfp-36; LIM metal-binding repeat homology
 C:Keywords: DNA binding; zinc finger

Query Match	16.8%; Score 374.5; DB 2; Length 710;
Beat Local Similarity	29.6%; Pred. No. 2.6e-15;
Matches	104; Conservative 54; Mismatches 140; Indels 53; Gaps 15;
QY	25 YICTYEGCDKAYRPSLLLEQHLRTHSNDPRYKCTVDDCDCAFPKSKLHETHTVSHSEKRP 84
DB	311 YEC--KECGKSFILSHLKHRYRHTHEKPKYCEI--CDKSFYTTTLTKTHQKHTHEKRP 366
QY	85 FHGVCVGKGVNSRQHLKRHEITHT--KSPKCTFENCOEAPYKQSLYHNLIS-VHEKTLT 141
DB	367 YKCBECDSKFLHSHLRRHQVHTGERPRYC--KECDKSHESATLEHEKSHTEKTYK 424
QY	142 CKQCNKYVTPRPSKLAQHKLKHGGSPAYOCSDHPGCFPNQFOTWSYLAQFN--IKOSHPRKLC 199
DB	425 CREDDKSFQRAVLYRNHNHVHTGERPYECKE--CGKSFCTCSTLRHQHTTHCKPKYC 482
QY	200 PKCGKCGVKKGLSSHMLSHDSTMIKIWTCDYCDVGKPAKQELVEHYNIIFHDGNI 259
DB	483 IECKGKFTNNSYLRTHQKHSIGE--KPYRCKECD-KSFSCSTLRKHQSI-HTGEKPKY 537
QY	260 LKRTYVKKLENLDDQSKLNNLHELETEKLYEDEDDEDDSLDEKRSVDSMSGAQR 319
DB	538 CI-----ECKDSFTQVSHLRTHQRYHTGERPRCTECD-----KSFIRSHSLREHQ 583
QY	320 SIKSFASLEGSKSVSKLINSGKK-INCPKNNCDRMFSREYDLRHLKMH 369
DB	584 KI-----HSGEKPKYC--RDCDISFSQISNLRHQKHL 614

RESULT 9
 A54661
 zinc finger protein ZNF41 - human (fragment)
 C:Species: Homo sapiens (man)
 C:Date: 02-Jun-1995 #sequence_revision 02-Jun-1995 #text_change 09-Jul-2004
 C:Accession: A54661; I54227
 R:Franchise, A.; Archidiacono, N.; Rocchi, M.; Marino, M.; Grimaldi, G.
 Genomice 9, 728-736, 1991
 A:Title: Isolation and expression analysis of a human zinc finger gene (ZNF41) located o
 A:Reference number: A54661; MUID:91244317; PMID:2037297
 A:Accession: A54661
 A:Status: Preliminary
 A:Molecule type: DNA
 A:Residues: 1-542 <PRA>
 A:Cross-references: UNIPROT:P51814; GB:M56548; GB:M92443; NID:g340443; PIDN:AAA61312.1;
 C:Gene: GDB:ZNF41
 A:Cross-references: GDB:125865; OMIM:314995
 A:Map position: Xp11.3-Xp11.23
 C:Superfamily: zinc finger protein zfp-36; LIM metal-binding repeat homology

Query Match	16.7%; Score 373; DB 2; Length 542;
Beat Local Similarity	29.0%; Pred. No. 2.4e-15;
Matches	115; Conservative 59; Mismatches 156; Indels 66; Gaps 21;
QY	14 ISSSSSRPKYITTYGCDKAYRPSLLLEQHLRTHSNDPRYKCTVDDCDKAPPKSHLE 73
DB	163 IHQKTHGEGKRYEC--NECKAFLFRKALMMHQRHTHEKPYVCA--DCGKAFIQKSHFN 218
QY	74 THYVSHSEKPPFHGVCVGKGVNSRQHLKRHEITHT--KSPKCTFENCOEAPYKQSLYH 131
DB	219 THQIHHTGEGKRYECSDGKAFYTKKSQALHQRHTHGKAPYICT--EGKAFYTHRTN---- 272
QY	132 ILVSHKTLT-----CKQCNKYVTPRPSKLAQHKLKHGGSPAYOCSDHPGCFPNQFOTWSY 185
DB	273 -LTHQKTHGEGKRYEACGKAFITDQSNLIKHQKHTGKRP-YKCN--GCGKAFIKSR 328
QY	186 LQPHIKOSH--PKLCPKPGKCGVCKGKGLSSHMLSHDSTMIKIWTCDYCDVGK-PAKK 241
DB	329 LKIH-QKSHGERHYECKDGKAFIQKSTSLVQRIH---TGKRPYVCPC--GKAFIQK 382

QY 242 NELVEVYNI FHDGNIPD-----LKETEYVKLTENLLDQSKNNLHE----- 284
 Db 383 SHPIAHRI-HTEBKPEYECSDCGCKFTKKSQI-RVHQKIHTEGKPNICAECKAFTDRN 440
 QY 285 -LETEKLTKEVEDEDEDS----LDEKRSVDYSDSMSAQRSI-----KSFTASLEGSKS 333
 Db 441 LITHQKIHREKYECCDCKFTFWKSRLLNHQSHGHERHYECSKCGAFIQ-KATLS 498
 QY 334 VSKLISNSGKINCPKNNCDRMFSREYDLRRHLKWH 369
 Db 499 MHQIHHTGKKPYAC--TECKAFTDRSNLHGHQKH 532
 RESULT 10
 B32891
 C:Species: Homo sapiens (man)
 C:Date: 20-Oct-1989 #sequence_revision 20-Oct-1989 #text_change 01-Dec-2000
 C:Accession: B32891
 R:Bellefroid, E.J.; Lecocq, P.J.; Benhida, A.; Poncelet, D.A.; Belayew, A.; Martial, J.A.
 DNA 8, 377-387, 1989
 A:Title: The human genome contains hundreds of genes coding for finger proteins of the K
 A:Reference number: A32891; MUID:89377476; PMID:2505992
 A:Accession: B32891
 A:Status: preliminary
 A:Molecule type: mRNA
 A:Residues: 1-651 <BEL>
 A:Cross-references: GB:M27878
 C:Superfamily: zinc finger protein ZFP-36; LIM metal-binding repeat homology
 C:Keywords: tandem repeat

```

Query March 16.7%; Score 373; DB 2; Length 651;
Best Local Similarity 27.4%; Pred. No. 2.9e-15;
Matches 128; Conservative 57; Mismatches 146; Indels 136; Gaps 24;

QY 23 KKIYICTYGCDAKYNRPSELTLEQHLRTHSNDRPYCTVDDCDKAFPRKSHLETHIVSH--- 79
Db 174 KPYNCQ---CGKAPSQKSLTSHQRTHTGKPYEC---GECGKAIRKSHLSHMTHTTGE 229
QY 80 -----SEK-----KPHCSVCGKVNRSQHLRHEITTH--TKSFK 112
Db 230 KPYGNECGRAFESEKSLINHQRIHTGKPEPCRECGKAFKRSQSLVTHHRTHTGTKPF 289
QY 113 CTFENCQCAF-KNQSLRHLHLSVHEKTLTKQCNKVPTRSKLAQNKLAHGGSP---- 167
Db 290 CS--DCRAKPEKSELINHQTIHTGKPEYSEBRKAFRRSSSLINHQRTHTGKPKGCI 347
QY 168 -----AVQCDHPG-----CPKNFQTSVLQFHLKOSHPKLK--CPKCG 203
Db 348 QCGKAFSQKSHLISHQWHTGKPEPICGCKGKASRSQSLVRH-QRHTHTGKPEPCGCG 406
QY 204 KGCYGGKGLSSHMLSHDSTMIKIWTCDYCDVGK-FAKQELVEHYNIFHDGIPDDL 261
Db 407 KAFSEKSLTINHQRH---TGEKPYVCSGC--GKAFQKSHLISHORT-HTGKPEPCSE 460
QY 262 -----KETEVKLEMLLDQGSGLNNLHLELFTBKLKVEDDEDE 300
Db 461 CGKAFGEKSLATQRTHTGKPEYECRCEPAFSQKSLNT-HQ-----RIHTGPEYE 513
QY 301 DSLDEKSDVSDSMSAQRISKSFASLEGSVSKSLINSNGK-INCPKNNDCRMFSRE 359
Db 514 CSLCRK-----AFPEKSELIRHLRTHHTGKPEYC--NECKAKREK 552
QY 360 YDLRRHLKMDNLIQRTSFLNSLEKETTPGGEPLVKKARMDLLPNE 406
Db 553 SSLINH-----QRHTGKPEPCSEC--GKAFRSKSH-LIPHQ 587

RESULT 11
finger protein pMLZ-4 - mouse
CISpecies: Mus musculus (house mouse)

```

C>Date: 31-Dec-1993 #sequence_revision 31-Dec-1993 #text_change 09-Jul-2004
C/Accession: JN0533
R:Brady, J. P.; Platigorsky, J.
Gene 124, 207-214, 1993
A>Title: Cloning and characterization of a novel zinc-finger protein encoding cDNA from
A/Reference number: JN0533, MUID:93185925, PMID:844344
A/Accession: JN0533
A/Molecule type: mRNA
A/Residues: 1-393

A/Cross-References: UNIPROT:003309, GB:M98502, NID:9200406, PTDN:AAA3949.1, PTD:G20040
A/Experimentally: zinc finger protein ZFP-36, LIM metal-binding repeat homology
C/Keywords: DNA binding; eye lens; zinc finger
P:1-62/Region: acidic
P:63-391/Region: zinc finger

Query Match 16.7%; Score 372; DB 2; Length 393;
Best Local Similarity 29.4%; Fred. No. 1.9e-15;
Matches 110; Conservative 57; Mismatches 163; Indels 44; Gaps 17;

Query Match	16.7%	Score 372	DB 2	Length 393
Best Local Similarity	29.4%	Pred. No. 1,9e-15		
Matches 110	Conservative 57	Mismatches 163	Indels 44	Gaps 17
QY	13	LISSSSSRPKKYITTEGCDKAYNRPBLLLEOHLTHNSDREPYKCTVDDCDCAFPRKSHL	72	
Db	49	LHNKAAHAGINHYHLSQ---CGKAFQSLDINRNHQHTHGDRYKYCY--ECGKGFSSSHL	104	
QY	73	ETHIVSHSEKKPRFHSCVCGKGVNSFOHLKRHEITTT--KSFCTENCOEAYYKQSL-R	129	
Db	105	IOHORHTHGERPPYDNECGKSPGRSSHIIOHOTHTGERPNKCT--ECKMAASPHLLQ	162	
QY	130	HHILSVHEKTLTLCOKCNVFTRPSTLAQHKLKHNGGSPAYOCDHPCFPNQFTWS--VHQ	187	
Db	163	HQHTHSGKPRYECBECGKSPFRSSHLAQHQRHTHEKPR-YEHE--CGRGEFERSLLIGH	219	
QY	188	FHIKOSHPLKCPKCGKGVCGKGLSSHMLSHDSTMIKIWTCDYCDVGEFAKKUELVEH	247	
Db	220	YRVHTGERPPYKCDCEGKFNQSDLDVRRRAH---TGEXPRYHCNCG--ENFSRISHLVQH	275	
QY	248	YNIPIHDGNIPPDLLKETEVKLENNLLDQSKLNNHLETEKLYKVEDE-----EBEEDS	302	
Db	276	QRT-HTGG-----PRYECTACGKSFSSSSSHITHQKIHGTGKPYECNECWRSFGRRSL	328	
QY	303	LDEKRSVRSMSMAQRSIKSFASLSESKSVSKLIS---NSGKK-INCPKNNCDRMP	357	
Db	329	IKHQRTHGEKPRYECVQCGKGFQTS-----SNLTIHQRVHTGKPRYEC--TECDKSTS	379	
QY	358	REYDRLRLKWHDD	371	
Db	380	RSSALIKKRRVYTD	393	

```

RESULT 12
G01496
transcription factor IIA - human (fragment)
N:Alternate names: TFIIA
C:Species: Homo sapiens (man)
C:Date: 21-Dec-1996 #sequence_revision 06-Jun-1997 #text_change 05-Nov-1999
C:Accession: G01496
R:Becker, K.G.
submitted to the EMBL Data Library, August 1994
A:Reference number: G07520
A:Accession: G01496
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1538 <BRC>
A:Cross-references: EMBL:U014134, NID:g551534, PIDN:AAA21873.1, PID:g551535
C:Genetics:
A:Gene: GDB:GTF3A; TFIIA
A:Cross-references: GDB:434744; OMIM:600860
A:Map position: 13q12.3-13q13.1
C:Superfamily: transcription factor IIA

Query Match          16.4%  Score 364.5;  DB 2;  Length 338;
Best Local Similarity 29.0%  Pred. No. 4.6e-15;
Matches    89;  Conservative 49;  Mismatches 102;  Indels    67;  Gaps    11;

```

```

OY      8 KSISL-----ISSSSSR-----PKYICYEECDXANRPSLLEQIATHSDR 53
Db      7 ESVSLLTIADAFIAGESSAPTPPPRALPRFICSPDCSANYSAMKLDMLCKHGER 66
OY      54 PYCTVDODCAFFPKSHLETHIVSHSEKDPHCSV--CGKGVNSRÖHLKEH-ETH--- 107
Db      67 PFCVCEGCGAFIFIDYHLSRHILITHTEKEPVCANCGDQKFNTKSLKGFEEKHENO 126
OY      108 TKSFCSTENCOEAEFYKHÖSLR-----HHILSVH 136
Db      127 ÖKÖYICSEDECKRTYKHÖQLKIHÖCÖHTNEBPLFCÖCKGCSFPAKTYTBLKHVRETH 186
OY      137 EKTLYTCÖKÖNVFTRPSYLAÖHKLKHGGSPAYÖCDBGCFKRNQÖTWSVLÖFHIKÖSHPK 196
Db      187 KEELICEVCRTKTFPKKDYLKÖHMKTHAPERDVYCHCPRECGRITVYVFNLOSHILSFEE 246
OY      197 LK---CPK--CGKGVCGKGLSSHMLSHD-DSTMIKIMTCÖYCDVCGFAKKNELVEHÖNI 250
Db      247 SRPFCEHAGCGKTFAMKÖSLTRHAVYHDPKCKMKL-----KYKGRERKRSILASHS- 299
OY      251 FHDGNIP 257
Db      300 ---GYIP 303

```

```

RESULT 13
A32891
finger protein 1, placental - human
C|Species: Homo sapiens (man)
C|Date: 20-Oct-1989 #sequence_revision 20-Oct-1989 #text_change 09-Jul-2004
C|Accession: A32891
R|Belletroid, E.J.; Lecocq, P.J.; Benhida, A.; Ponclet, D.A.; Belayew, A.; Martial, J.F.
DNA 8, 377-387, 1989
A|Title: The human genome contains hundreds of genes coding for finger proteins of the R
A|Reference number: A32891; MUID:89377476; PMID:2505992
A|Accession: A32891
A|Status: preliminary
A|Molecule type: mRNA
A|Residues: 1-428 <BL>
A|Cross-references: UNIPROT:P51522; GB:M27877
C|Superfamily: zinc finger protein ZFP-36; LIM metal-binding repeat homology
;Keywords: tandem repeat

```

Query Match	16.3%	Score 364	DB 2	Length 428
Best Local Similarity	29.1%	Pred. No. 6,4e-15		
Matches	116	Conservative 50	Mismatches 155	Indels 78
				Gaps 18
QY	32	CDKAVNRSLLEQHRTHSNDRPYKCTVDDCDKAFRRSHLETHVSHSEKKPFHSCVG	91	
Db	38	CGKLFNKKSNLASHORLHTGEKPYKC--NECGKVFHNNSHLAQHRRIHTGEKPYKCNECG	95	
QY	92	KGAVNRQLKRRHEITHT--KSPKCTFENCQGLAPYKHSL--RHHILSVHEKTLTLTKQCNKV	148	
Db	96	KVFNQISHLAQRHRTHTGEKPYKC--NECGKVFHQISHLAQRHTHTGEKPYECNCKGKV	153	
QY	149	FTPRSKLAQHKLKHIGGSPAYOCDHQCGCFKQFQWTSVLYQFL--IKQSHPLKCDPKCGKC	206	
Db	154	FSRNSYLVQHLIHTGEKPYRCN--VCGKVFHNI SHLAQHRHTHTGEKPYKCNECGKVF	210	
QY	207	VGKKGLSGHMLSHDSTWIKIWTCDYCDVYG--PAKKNELVHYNVPHQGNIP-----	257	
Db	211	SHKSLVNNHMRH--TGEKPYKCNEC--GKVFHSKSLVNNHMR--HTGEKPYKCNECGK	264	
QY	258	-----DDL-----KETEVKKLENTLQOGSKLNNLHETLEKTLVEDEDEDEDSL	303	
Db	265	VFSRNSYLAQHLIHTAGEKPYKCDECDKAFBSQNSHLVQHRHTHTGEKPYKCDGCGKFSQ	324	
QY	304	D-----EKTSQDVR-----SDPSMAQRST-----KSFYASLEG	330	
Db	325	NSYLAHYHRHTHTGEKPYKCNECGKVFGLNSSLAHRKHTHTGEKPYKCNECGKAF--SWRS	382	
QY	331	SKSVSKLISNGKINCPCNNCDRMFSREYDLRRLHLKWH	369	

```

Db          363 SLTWHAIHTGKKFKC--NECGTLFRDNSYLVRHQRFH 419
               | | | | | | | | | | | | | | | | | |
RESULT 14
506546
finger protein (clone X1COP7.1) - African clawed frog (fragment)
C|Species: Xenopus laevis (African clawed frog)
C|Date: 28-Feb-1990 #sequence_revision 28-Feb-1990 #next_change 09-Jul-2004
C|Accession: S06546
R|N|E|F|e|l|d|_W|_B|_B|a|r|a|t|_T|_M|e|n|t|z|e|l|_H|_P|i|e|t|e|r|_T|_K|o|e|s|t|e|r|_M|_P|o|e|t|i|n|g|_A|_K|n|o|e|c|h|
J|_M|o|l|_B|io|l|_2|0|8|_6|3|9|-6|5|9|_1|9|8|9
A|Title: Second-order repeats in Xenopus laevis finger proteins.
A|Reference number: S05632; MUID:90040698; PMID:2509712
A|Accession: S06546
A|Status: not compared with conceptual translation
A|Molecule type: mRNA
A|Residues: 1-615 <NIE>
C|Cross-references: UNIPROT:P18751
C|Superfamily: zinc finger protein ZFP-36; LIM metal-binding repeat homology
C|Keywords: DNA binding; zinc finger

```

Query Match	16.3%;	Score 363.5;	DB 2;	Length 615;
Best Local Similarity	27.6%;	Pred. No. 1e-14;		
Matches 119;	Conservative 48;	Mismatches 165;	Indels 99;	Gaps 18

OY	1	SSLSSSSSSSRPKYKITYEGCDKAYNRPSSLEQLRTHSNRPYKCTYDDCDKAFRKS	70
Db	128	SSLVNHQRTHTIGKEPFCSE--CDKCFASSSELNIHQRTHTGKPFSCS--EGCKPTNHS	184
OY	71	HLEETIVSHSEKKRPHCSVCVGKGVNSROHLKHEITHT--KSPKCTFENCQOAFYHQSL	128
Db	185	HPAHNHQMTHTGKEPFCSCCKGKCFASSSDLTGHRHTHREKTFSCS--ECGKCFSHSHL	242
OY	129	-----RH	130
Db	243	ARHQMHTGKEPFCSCSECGKCFSSSSGLTAHQRTHMKVPCSCAGKCFSNRSHLHH	302
OY	131	HILSVHEKTLCKOCNKVFTPRPSKLAOKHLKHGSSPAVOCNPGCFKNPOTWSVLOPHI	190
Db	303	QMHTGKEPFCSECFERKCFNSPNSLARHQMTTGKEP-FCSE--CGKCFASSSDLTGH-	358
OY	191	KOSHPKLT--CPKCGKGVGK--GLSSHMLSHDSTMIKITYCDYGVGKPAKNELVE	246
Db	359	HRHTGKEPFCSECGK-CYSKSSSLVNHQRTHTGKEPFCSCDKC-----PASSSEL--	411
OY	247	HYNI FHDONI DDLLKETEYVKKLEVLDOGSKLTNNLHLETFEKLAVEDBEDEBDSLDER	306
Db	412	--NIHQRTHEKAPFCSECGKCFPTNRSQULSHQMIHTGE-PRICPBECECFVSS----	464
OY	307	RSVDKSDMSAORSISFTASLEGSKVS-----KLISNGKIKINCPRKNOCBDFGR	358
Db	465	-SGLTAAHQOQAHRWVKPFC--LECKCFCSNNSNFPARHQMTHTGKEPFC--SECRKGFEN	520
OY	359	EYDLRRHLKMH	369
Db	521	QSSLARHQMTH	531

```

RESULT 15
A35659
knueppel-related protein H-plk - human
C:Species: Homo sapiens (man)
C:Date: 28-Sep-1990 #sequence_revision 18-Nov-1992 #text_change 09-Jul-2004
C:Accession: A35659
R:Kato, N.; Shimotohno, K.; VanLeeuwen, D.; Cohen, M.
Mol. Cell. Biol. 10, 4401-4405, 1990
A:Title: Human prostatic mRNAs down regulated in choriocarcinoma encode a zinc finger pro
A:Reference number: A35659; MUID:90318410; PMID:2115127
A:Accession: A35659
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-427 <KAT>

```

A;Cross-references: UniProt:Q02313; GB:M55422; NID:gl184342; PIDD:AAA6010.1; PID:gl184343
C;Superfamily: zinc finger protein ZFP-36; LIM metal-binding repeat homology

Query Match 16.1%; Score 359.5; DB 2; Length 427;
Best Local Similarity 33.0%; Prod No 1 30-14.

```

QY 20 SRRKTKICYEGGDKAYNRPSSLEQHLNTHSNDRYCTVDDCCAKFRRSHLETHVSH 79
Dh 132 TRNRFKCC--EAYGRAFNWSTLTHKKIHTGEKPYKC--KECGAFNQSHTLRRHRH 187
QY 80 SEKKRPCHSVCGKVNRSQHLEKHEIYTHKS--FKCTPENCOBAFYKHOSL-RPHILSVH 136
Dh 188 TEKKRPKECECGKAFNQSSTLTTHNIHTGELPYKC--EKCVRAFNQASKTLEKHLHTG 245
QY 137 EKTLTCKQCNKVFTREPSKLAQHKLH-----HGSPAY 169
Dh 246 EKRYECECGKAFNRSSKLTENKIHTGEKLYKCEBGKAFNQSSTLTTHKRIHSGEKPY 305
QY 170 QCDHPCGCFKFNQFTWSYLQFH--IKOSHPLKCPKCGKCGVKGKGLSSHMLSHDSTWIKI 227
Dh 306 KCEE--CGRAFKQFNSLTDHKKIHTGEKPYKCEBGKAFNQLSNLTBRKVIH--TGEEK 360
QY 228 WTCDDYDVKRPAKKMEIVENYNIHFDGNIHPPDLLKETPVKLEMLDQGSKLANLHELET 287
Dh 361 YKCGEC--SKAFNQSALNTHKIHTGENPHKCRSGEVFHL-----SSKLTCKYKHT 412
QY 288 -EKL-KVEE 294
Dh 413 GEKLYKCEE 421

```

Search completed: February 8, 2005, 15:22:57
Job time : 51 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 8, 2005, 15:11:59 ; Search time 179 Seconds
(without alignments)
1178.640 Million cell updates/sec

Title: US-09-831-804-3
Perfect score: 2229
Sequence: 1 MSBDETKSISLSISSSSSS.....PLYKARMDLLPNETYSISR 412

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1612378 seqs, 512079187 residues

Total number of hits satisfying chosen parameters: 1612378

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : UniProt 03:.*
1: uniprot_sprot:.*
2: uniprot_trembl:.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1257	56.4	452	2	06BU9
2	850	38.1	472	2	06CK1
3	837.5	37.6	429	1	TF3A.YEAST
4	819	36.7	476	2	Q758N3
5	809.5	36.3	394	2	Q6FS08
6	495	22.2	374	1	TF3A.SCHPO
7	470.5	21.1	655	2	06C2D4
8	411.5	18.5	634	2	06R5P3
9	410.5	18.4	634	2	06R5P4
10	410.5	18.4	634	2	06B1B5
11	407	18.3	611	2	Q7RY67
12	400	17.9	404	2	08N7O3
13	400	17.9	803	1	ZN43.HUMAN
14	398	17.9	574	1	Z492.HUMAN
15	397.5	17.8	395	2	Q436Y3
16	397.5	17.8	423	1	TF3A.HUMAN
17	396.5	17.8	595	1	ZN85.HUMAN
18	395	17.7	616	1	ZN93.HUMAN
19	393.5	17.7	1191	1	ZN91.HUMAN
20	392.5	17.6	542	2	081YV0
21	392	17.6	553	2	08N5N8
22	392	17.6	644	2	08ND40
23	392	17.6	661	2	Q8B6W5
24	391	17.5	592	2	Q8N8O4
25	390	17.5	540	2	Q8N211
26	390	17.5	576	2	Q96N22
27	389	17.5	891	2	Q6ZNA1
28	388.5	17.4	501	2	Q9VG72
29	388	17.4	568	2	Q8TD23
30	387.5	17.4	839	1	Z347.HUMAN
31	385.5	17.3	297	2	Q8B0T6

32	385	17.3	576	1	Z431.HUMAN	Q8tf32	homo sapien
33	384	17.2	753	2	Q6P8I0	Q6P8I0	mus musculus
34	383	17.2	913	1	Z228.HUMAN	Q9Uj33	homo sapien
35	382	17.1	511	2	Q8VRC1	Q8VRC1	mus musculus
36	381	17.1	493	2	Q92IH6	Q92IH6	mus musculus
37	379.5	17.0	710	2	Q91VP4	Q91VP4	mus musculus
38	379	17.0	600	2	Q8CE00	Q8CE00	mus musculus
39	379	17.0	766	2	Q8CAJ2	Q8CAJ2	mus musculus
40	378.5	17.0	914	1	ZF62.MOUSE	Q8C827	mus musculus
41	378	17.0	393	2	Q923E2	Q923E2	mus musculus
42	378	17.0	452	2	Q8BPP0	Q8BPP0	mus musculus
43	378	17.0	452	2	Q8BQC2	Q8BQC2	mus musculus
44	378	17.0	461	2	Q69ZC6	Q69ZC6	mus musculus
45	378	17.0	613	2	Q9H7R5	Q9H7R5	homo sapien

ALIGNMENTS

RESULT 1

ID	Q6BU9	PRELIMINARY;	PRT;	452 AA.
AC	Q6BU9;			
DT	25-OCT-2004 (TrEMBLrel. 28, Created)			
DT	25-OCT-2004 (TrEMBLrel. 28, Last sequence update)			
DT	25-OCT-2004 (TrEMBLrel. 28, Last annotation update)			
DE	Similar to sp P39933 Saccharomyces cerevisiae YPR186c TRC2			
DE	TFIIIA.			
GN	ORFNames=DEHA0C07623g;			
OS	Debaromyces hanseni CBS6767.			
OC	Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;			
OC	Saccharomycetales; Saccharomycetaceae; Debaromyces.			
OX	NCBI_TaxID=284592;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=CBST67;			
RG	Genolevures;			
RA	Dujon B., Sherman D., Fischer G., Durrens P., Casaregola S., Lafontaine I., de Montigny J., Marck C., Neveglisse C., Talia E., Goffard N., Frangeul L., Algile M., Anthouard V., Babour A., Barbe V., Barnay S., Blanchin S., Beckertich J.M., Beyne E., Bleykasten C., Boistrame A., Boyer J., Cattolico L., Confaniolero F., de Darvar A., Despons L., Fabre B., Fairhead C., Ferry-Dumazet H., Groppi A., Hantraye F., Hennequin C., Janniaux N., Joyet P., Kachouri R., Karrest A., Kozul R., Lemaitre M., Loeur I., Ma L., Muller H., Nicaud J.M., Nikolski M., Oztas S., Oztur-Kalogeropoulos O., Pellenc S., Potier S., Richard G.F., Straub M.L., Suleau A., Swemene D., Tekala F., Wesolowski-Louvel M., Westhof E., Wirth B., Zeniou-Meyer M., Zivanovic I., Bolotin-Fukuhara M., Thierry A., Bouchier C., Caudron B., Scarpelli C., Galliard C., Weisenbach J., Wincker P., Souciet J.L.;			
RT	"Genome evolution in yeasts.";			
RL	Nature 430:35-44(2004).			
RN	[2]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=CBST67;			
RA	Genoscope;			
RL	Submitted (JUL-2004) to the EMBL/GenBank/DBJ databases.			
DR	EMBL; CR82135; CAG86038.1; -.			
DR	GO; GO:0005634; C:nucleus; IEA.			
DR	GO; GO:0003676; F:nucleic acid binding; IEA.			
DR	GO; GO:0008270; F:zinc ion binding; IEA.			
DR	InterPro; IPR007087; Znf_C2H2.			
DR	Pfam; PF00096; Zf_C2H2_8.			
DR	SMART; SM00355; Znf_C2H2_9.			
DR	PROSITE; PS00028; ZINC_FINGER_C2H2_1; 8.			
DR	PROSITE; PSS0157; ZINC_FINGER_C2H2_2; 7.			
SO	SEQUENCE 452 AA; 52049 MW; D8224533C04B0269 CRC64;			

Query Match 56.4%; Score 1257; DB 2; Length 452;
Best Local Similarity 55.6%; Pred. No. 2,5e-64;
Matches 234; Conservative 62; Mismatches 93; Indels 32; Gaps 10;

```

QY 3 ESDETSSISSSSSSRPKKYICTYEGCDKAYNRPSSLLEQHLKTHSNDRPKYCTVDDC 62
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 2 ESEETASVSR-ASSSSARPKNTICBPDGKAYSPSLLEQKRRHTNRPKCSGTGC 60
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 63 DKAFPRKSHLETHIVSH--SEKKPFCQSVCGKGVNSRQHLKREHITHTYSFKCTFENCQ 120
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 61 DKSFLKSHLQAHLLSHBEDSKPFCQCATGKGVNTLQHLKREHITHTYSFICTFEGCN 120
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 121 APTKHOSLRHHILSVHEKTLTKQCKNVFTRPCKLAQHLKHHGSDPAVQCDHPGCPKN 180
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 121 SPYKHOSLRHHITLSVHEKTLCKNKCKNSFNRPYRLAQHNIKYHSDSPAYQCDHQGCGNF 180
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 181 QTMVSVOFHIKOSHPRKLCCKPCGKGVCKKGLSSHMLSHDSTMTIKITCDYCDYCKFAK 240
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 181 MTMSAQLQHLKTHPRKICPCTCGKGVCKKGLSSHMLSHDSTMTIKITCDYCDYCKFAK 240
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 241 KNELVHYNIFHDGNTPDLLKTEVKKLENLDQGSK----LNNHLETERK-LKVEED 295
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 241 KADLIHNTYHNRNLPDDLKPIREQLDKLSHSDSTNIDPNTLEGLOSKGFVPSD 300
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 296 EEDDEDSLDEKRSVDVSDMSA-QRSIKSFTASLE-GSKSVSKLISNS--GKKINCPKN 351
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 301 EEKKE-----NDYMSNGLAQSHKSLNSLNTLESGKASIVDLLNNYLKRVPCPKK 353
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 352 CDMFEREYDLRHLKWHNDNLORISPLNSIKEET-----PEG-EPVYKA 398
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 354 CDMFERDYDLERHLKWHESHLLKRIAPFLKSLLEKVPDSNMHDKRLPEPDEPWATK 413
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 399 R 399
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 414 R 414
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

```

RESULT 2

```

OCWK1 PRELIMINARY; PRT; 472 AA.
ID 06CWL;
AC 06CWL;
DT 25-OCT-2004 (TREMBlrel. 28, Created)
DT 25-OCT-2004 (TREMBlrel. 28, Last sequence update)
DT 25-OCT-2004 (TREMBlrel. 28, Last annotation update)
DE Similarities with sp|P39933 Saccharomyces cerevisiae YPR166C TFC2
DE TFI1A.
GN ORFNames=KLIAB03454g;
OS Kluyveromyces lactis NRRL Y-1140.
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Kluyveromyces.
OX NCBI_TaxID=284590;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=NRRL Y-1140;
RG Genolevures;
RA Dujon B., Sherman D., Fischer G., Durrens P., Casaregola S.,
RA Lafontaine I., de Montigny J., March C., Neuvéglise A., Talla E.,
RA Coffard N., Frangeul U., Aigle M., Anthouard V., Babour A., Barbe V.,
RA Barnay S., Blanchin S., Beckerich J.M., Beyne E., Bleykasten C.,
RA Boissame A., Boyer J., Cattoilico L., Confanioleri F., de Daruvar A.,
RA Despons L., Fabre E., Fairhead C., Ferry-Dumazet H., Gropi A.,
RA Hantaye F., Henneguin C., Janniaux N., Joyet P., Kachouri R.,
RA Kerrast A., Koszul R., Lemaire M., Leort I., Ma L., Muller H.,
RA Nicard J.M., Nikolski M., Oztas S., Oztier-Kalogretopoulos O.,
RA Pellenz S., Potier S., Richard G.F., Straub M.L., Suleau A.,
RA Swenneu D., Tekala F., Wesolowski-Louvel M., Westhof E., Wirth B.,
RA Zenitun-Meyer M., Zivanovic I., Bolotin-Fukuhara M., Thierry A.,
RA Bouchier C., Caudron B., Scarpelli C., Galliardin C., Weisenbach J.,
RA Wincker P., Souciet J.L.
RT "Genome evolution in yeasts.";
RL Nature 430:35-44(2004).
[2]
RP SEQUENCE FROM N.A.
RC STRAIN=NRRL Y-1140;
RG Genoscope;
RT Submitted (JUL-2004) to the EMBL/GenBank/DBJ databases.
RL EMBL; CR382122; CAH02081.1; -.

```

```

DR GO; GO:0005634; C:nucleus; IEA.
DR GO; GO:0003676; F:nucleic acid binding; IEA.
DR GO; GO:0008270; F:zinc ion binding; IEA.
DR InterPro; IPR07087; Znf_C2H2.
DR Pfam; PF00096; zf-C2H2; 8.
DR SMART; SM00355; ZNF_C2H2; 9.
DR PROSITE; PS00028; ZINC_FINGER_C2H2_1; 8.
DR PROSITE; PS00157; ZINC_FINGER_C2H2_2; 5.
SQ SEQUENCE 472 AA, 54229 MW, 5AB712DB2674A4D CRC64;

```

Query Match 38.1%; Score 850; DB 2; Length 472;
 Best local similarity 45.3%; Pred. No. 5 6e-41;
 Matches 189; Conservative 48; Mismatches 128; Indels 28; Gaps 7;

```

QY 9 SSSSSSSSSSSRPKKYICTYEGCDKAYNRPSSLLEQHLKTHSNDRPKYCTVDDCDAKF 67
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 90 STSTSVSVSSQSRKVFCDYDQCKNSFTRPRLTLEHQLTVHNGIKPFC--DTGCKERA 147
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 68 RKSHEETHIVSHSEKKPFCQSVCGKGVNSRQHLKREHITHTYSFKCTFENCQAPKQ 127
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 148 KXSHNRHMFSTIDKPFQCSICGKGVYTRQOLKREHITHTYSFQSVYGCNESPYPK 207
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 128 LRHHILSVHEKTLTKQCKNVFTRPCKLAQHLKHHG--GSPAYOCDHPGCFKNPQTSV 185
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 208 LRSHILSVHEKTLTKPCNKCTFQRPYRLKNHIDKHNBSGTMYQCDPLSTDVFSWSS 267
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 186 LQFHIKOSHPRKLCCKPCGKGVCKKGLSSHMLSHDSTMTIKITCDYCDYCKFAKKNELV 245
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 268 LQGHIKQCHPRKLCPCICGKGVASGLRNHMHIBESLVTKMKKSSCPDTSFAKKTQV 327
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 246 EHYNIFHDGNTPDLLKTEVKKLENLDQGSKLNHLETERK-LKVEEED----- 298
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 328 QHYESHKQVDEPDLLELPTDADP-----DTNKKVNT-----DKISTEBSFSGKKT 375
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 299 --EEDSLDEKRSVDVSDMSAQRSIKSFTASLESGKSVSKLISNSGKKINCPKNCDPMF 356
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 376 LIDPSPFSKRVDELGSVEAEIKIKYTES--GKTTSLNLTWQGRKCPYTNCKRFT 433
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 357 SREYDLRRLHKWH 369
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 434 KSEKEFDTHIEKH 446
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

```

RESULT 3

```

TF3A_YEAST STANDARD; PRT; 429 AA.
ID P39933;
AC P39933;
DT 01-FEB-1995 (rel. 31, Created)
DT 01-FEB-1995 (rel. 31, Last sequence update)
DT 05-JUL-2004 (rel. 44, Last annotation update)
DE Transcription factor TFI1A (TFI1A).
GN Name=TFI1A; Synonyms=BZFL, TFI1A; OrderedLocustNames=YPR166C;
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92147684; PubMed=1737784;
RA Archambault J., Milne C.A., Schappert K.T., Baum B., Friesen J.D.,
RA Segall J.E.
RT "The deduced sequence of the transcription factor TFI1A from
RT Saccharomyces cerevisiae reveals extensive divergence from Xenopus
RT TFI1A."
RL J. Biol. Chem. 267:3282-3288(1992).
[2]
RP SEQUENCE FROM N.A.
RX MEDLINE=92237295; PubMed=1570325;
RA Woychik N.A., Young R.A.;
RT "Genes encoding transcription factor TFI1A and the RNA polymerase
RT common subunit RPB6 are divergently transcribed in Saccharomyces
RT cerevisiae."

```

RL Proc. Natl. Acad. Sci. U.S.A. 89:3999-4003(1992).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=5288c / AB972;
RX MEDLINE=97313271; PubMed=9169875;
RA Bussey H., Storms R.K., Ahmed A., Albermann K., Allen E., Ansoorge W.,
RA Araujo R., Aparicio A., Barrell B.G., Badcock K., Benes V.,
RA Botstein D., Bowman S., Bruckner M., Carpenter J., Cherry J.M.,
RA Chung E., Churcher C.M., Coster F., Davis K., Davis R.W.,
RA Dietrich F.S., Delius H., DiPaolo T., Dubois E., Duesterhoeft A.,
RA Duncan M., Floeth M., Fortin N., Friesen J.D., Fritz C., Goffeau A.,
RA Hall J., Heeling U., Heumann K., Hilbert H., Hillier L.W.,
RA Hunkeler-Smith S., Hyman R., Johnston M., Kalkan S., Kleine K.,
RA Komp C., Kundi O., Laebkari D., Lew H., Lin A., Lin D., Louis E.J.,
RA Marabe H., Messenguy F., Mewes H.-W., Mitternacht S., Moestl D.,
RA Mueller-Auer S., Namath A., Netuschil U., Oefner P., Pearson D.,
RA Petel F.X., Pohl T.M., Purnelle D., Schaefer M., Scharfe M.,
RA Scherens B., Schramm S., Schroeder M., Sidcu A.M., Tettein H.,
RA Urrutazu L.A., Ushinsky S., Vierendeels F., Vissers S., Voss H.,
RA Walsh S.V., Wambut R., Wang Y., Wedler E., Wedler H., Winnett E.,
RA Zhong W.W., Zollner A., Vo D.H., Han J.,
RT "The nucleotide sequence of Saccharomyces cerevisiae chromosome XVI.",
RL Nature 367:103-105(1997).
CC -I- FUNCTION: Interacts with the internal control region (ICR) of
CC approximately 50 bases within the 5S RNA genes, is required for
CC correct transcription of these genes by RNA polymerase III. Also
CC binds the transcribed 5S RNA's.
CC -I- SUBCELLULAR LOCATION: Nuclear.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: M80611; AAB08014.1; -;
DR EMBL: M90638; -; NOT ANNOTATED_CDS.
DR EMBL: U25841; AAB64615.1; -;
DR PIR: S20050; S20050.
DR HSSP: P25490; IUBD.
DR Germonline: 144451; -;
DR TRANSFAC: T03530; -;
DR SGD: S00006390; PZFL.
DR GO: GO:0005677; C:transcription factor complex; IPI.
DR GO: GO:0003709; F:RNA polymerase III transcription factor act. .; IPI.
DR GO: GO:0006384; P:transcription initiation from Pol III promoter; IPI.
DR InterPro: IPR007087; Znf C2H2.
DR Pfam: PF00096; zfc-C2H2; 5.
DR SMART: SM00355; ZNF_C2H2; 9.
DR PROSITE: PS00028; ZINC_FINGER_C2H2_1; 8.
DR PROSITE: PS0157; ZINC_FINGER_C2H2_2; 6.
DR DNA-binding; Metal-binding; Nuclear protein; Repeat; RNA-binding;
KW Transcription regulation; Zinc-finger.
FT DOMAIN 24 43 Ser-rich.
FT ZN_FING 49 74 C2H2-type.
FT ZN_FING 80 102 C2H2-type.
FT ZN_FING 108 130 C2H2-type.
FT ZN_FING 134 159 C2H2-type.
FT ZN_FING 163 186 C2H2-type.
FT ZN_FING 194 219 C2H2-type.
FT ZN_FING 222 244 C2H2-type.
FT ZN_FING 253 277 C2H2-type.
FT ZN_FING 321 327 Arg/Lys-rich (basic).
FT DOMAIN 365 389 C2H2-type.
FT ZN_FING 429 449 C2H2-type.
SQ SEQUENCE 429 AA; 50027 MW; 209B1EDBA20422D9 CRC64;
Query Match 37.6%; Score 837.5; DB 1; Length 429;
Best Local Similarity 43.7%; Pred. No. 2; Ce-40;
Matches 174; Conservative 62; Mismatches 135; Indels 27; Gaps 11;
QY 1 MSESDETKSISSLS--SSSSSRPKYICTYEGCDKAYNRPISLLEQH-LATHSNDPRYKC 57

Db 23 ISRSESSESLNLTSTRSSSSNRPKTYPCDYDGCDAFTRPSTLTHQSLVHGLAFC 82
QY 58 TYDDCKAFPKSHLETHIVSHSEKPPHCSVCGKAVNSRHLKREIHTTSGFKCTFEN 117
Db 83 --DKCAKSPVKSHLERHLVTHSDTRPFCQSYCGKAVTTRQOLKRHEVTHTGSFICEEG 140
QY 118 COEAFKQDLSLRHLISVYEKTLTCKQCNKVFTRPEKLAQHLKHGG--SPAYQCDHPG 175
Db 141 CNLRFYKHQQLRHLISVHLHLKLTCPHCKNSFQRPRLRNHLSKHDPEVENYQCTFAG 200
QY 176 CFKNPQTVSVLPFHINQSHPKLCPKCGKGVCKGLSSHMLSHDSTWIKITWTCYCDV 225
Db 201 CCKEPFHWQLOSHIINDPKLCPICSKPCVGENGLQNMHIHDDSLVTKWKKHICPD 260
QY 236 GKPAKNEIYEVNINR-DGNTPDD-LKTEVYKLENLDDQSKLNLHELETYKAYE 293
Db 261 MFSRKHDLITHTGSHTEEDIPLKLYKISDQOL--VQDHGVQDGN-----SKHSNE 312
QY 294 EDEDEDESDLEKRSQVSDSMSAORSIKSFASLE-GSKSVKLSNSGKKINCPKNC 352
Db 313 QDEKTSNLRKRKRKLTENNVEFLQNEVDLEKRLSEGENGLNLNTVGRKTRCYNNC 372
QY 353 DRMFREYDLRRL--KMHDDNLQRISEFLNSIEKEE 387
Db 373 SRTFKTKKYEKHIIDGKVELKILQ-----EKEE 404
RESULT 4
QY 0758N3 PRELIMINARY; PRT; 476 AA.
AC 0758N3;
DT 05-JUL-2004 (TREMblrel. 27, Created)
DT 05-JUL-2004 (TREMblrel. 27, Last sequence update)
DT 05-JUL-2004 (TREMblrel. 27, Last annotation update)
DE AEL278WP.
GN ORFNames=AEL278W;
OS Ashbya gossypii (Yeast) (Eremothecium gossypii).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Eremothecium.
OX NCBI_TaxID=33169;
RN [1]
RP SEQUENCE FROM N.A.
RA STRAIN=ATCC 10895;
RA Gates K., Dietrich F.S., Brachat S., Voegelé S.E., Lerch A.,
RA Phillipsen P., Gaffney T.;
RL Submitted (SEP-2004) to the EMBL/GenBank/DBJ databases.
DR EMBL: A5016818; AAS52406.1; -;
DR HSSP: P03001; IUN6.
DR ACD; AEL278W; -;
DR GO: GO:0005634; C:nucleus; IEA.
DR GO: GO:0003676; F:nucleic acid binding; IEA.
DR GO: GO:0008270; F:zinc ion binding; IEA.
DR InterPro: IPR007087; Znf C2H2.
DR Pfam: PF00096; zfc-C2H2; 8.
DR SMART: SM00355; ZNF_C2H2; 9.
DR PROSITE: PS00028; ZINC_FINGER_C2H2_1; 8.
DR PROSITE: PS0157; ZINC_FINGER_C2H2_2; 3.
DR PROSITE: PS0157; ZINC_FINGER_C2H2_2; 3.
SQ SEQUENCE 476 AA; 53433 MW; D1467545D4E6BE0D CRC64;
Query Match 36.7%; Score 819; DB 2; Length 476;
Best Local Similarity 40.8%; Pred. No. 3; Ae-39;
Matches 165; Conservative 65; Mismatches 136; Indels 38; Gaps 9;
QY 2 MSESDETKSISSLS--SSSSSRPKYICTYEGCDKAYNRPISLLEQHLT-HSNDPRYKTV 60
Db 96 TSAGSTVSLSSVSTASTSTRKTHVCDYEGCYKAFTRPISLLEHQTAAQGRAQCC--E 153
QY 61 DCDKAFPKSHLETHIVSHSEKPPHCSVCGKAVNSRHLKREIHTTSGFKCTFENCOE 120
Db 154 QCGRGFTKSHLERHLVSHSEKPPHCSVCGKAVTTRQOLKRHEIHTTSGFKCPHGGCB 213
QY 121 AFYKHQSLRHLISVYEKTLTCKQCNKVFTRPEKLAQHLKHGGSPA-----YQCDHPG 176

Db 214 AFKPHQLSHVLAHVHQKLTCTHCKRQRPRLTLTHAKHHG--PAQOFRQCTNAC 271
Qy 177 FKNFQTSVLOPHIKOSHPLKCPKCGCVGKGLSHMLSHDSTMIKIWTCDYCVG 236
Db 272 VQCFETWSALQOLHLDHRLPCGVGCKLCVGETGQIMTVHDESRVLIKMKCSVCSPT 331
Qy 237 KFAKNEIYVHNIPHDGNIIPDLLKETEYKLENLDDQSKLNNLHELETEKLYEBDE 296
Db 332 TYAKMADLLAHYKTKGDSIPKELIHHVNAE-----SPAVHYVDEQQ 373
Qy 297 EDESDLDERSVDRD-----SMSAQRISKFTASLEGSKSVKLSNGKINPKNN 351
Db 374 REDPELVPRQKRKNSDNTIMGSLQTEEKIRKIES--GRTGSLTLNATGRRKRCPYLG 431
Qy 352 CDRMFGR--YDLRHLKWHDDNLQRIESFLNSIEKEETPECEP 393
Db 432 CSRVFTEDEKYD--HISKHINDLKVLLKEDLTKTEAAGNDP 473

RESULT 5

Q6FSJ8 PRELIMINARY; PRT; 394 AA.
AC 06FSJ8;
DT 05-JUL-2004 (T-EMBLrel. 27, Created)
DT 05-JUL-2004 (T-EMBLrel. 27, Last sequence update)
DT 05-JUL-2004 (T-EMBLrel. 27, Last annotation update)
DE Similar to sp|P39933 Saccharomyces cerevisiae YPR18c TFC2.
GN ORFNames=CG10610021g;
OS Candida glabrata CBS138.
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; mitosporic Saccharomycetales; Candida.
OX NCBI_TaxID=284593;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CBS138;
RG Genolevures;
RA Dujon B., Sherman D., Fischer G., Durrens P., Casaregola S.,
RA Lafontaine I., de Montigny J., Marck C., Neveglisse C., Talla E.,
RA Goffard N., Fringault L., Aigle M., Anthonard V., Babou A., Barbe V.,
RA Barnay S., Blanchin S., Beckerich J.M., Bayne E., Bleykasten C.,
RA Boistrame A., Boyer J., Catolico L., Confenietti F., de Daruvar A.,
RA Despons L., Fabre E., Fairhead C., Ferry-Dumazet F., Gropi A.,
RA Hantaye F., Hennequin C., Jaunhaux N., Joyet P., Kachouri R.,
RA Kerest A., Koszul R., Lemaire M., Leduc I., Ma L., Muller H.,
RA Nicaud J.M., Nkolisi M., Oztas S., Oster-Kalogeropoulos O.,
RA Pellenz S., Potier S., Richard G.F., Straub M.L., Suleau A.,
RA Swemene D., Tekala F., Wesolowski-Louvel M., Westhof E., Wirth B.,
RA Zenlou-Meyer M., Zivanovic I., Bolotin-Fukuhara M., Thierry A.,
RA Bouchier C., Caudron B., Scarpelli C., Galliardin C., Weissenbach J.,
RA Wincker P., Souciet J.L.;
RT "Genome evolution in Yeasts."
RL Nature 430:35-44(2004)
DR EMBL/CR380953; CAG59723.1; -;
DR GO; GO:0005634; C:nucleus; IEA.
DR GO; GO:0003676; F:nucleic acid binding; IEA.
DR GO; GO:0008270; F:zinc ion binding; IEA.
DR InterPro; IPR007087; Znf_C2H2.
DR Pfam; PF00096; zf-C2H2; 8.
DR SMART; SM00355; ZNF_C2H2; 9.
DR PROSITE; PS00028; ZINC_FINGER_C2H2_1; 8.
DR PROSITE; PS50157; ZINC_FINGER_C2H2_2; 4.
DR PROSITE; PS50157; ZINC_FINGER_C2H2_2; 4.
SQ SEQUENCE 394 AA; 45716 MW; 130B61605EFC799 CRC64;

Query Match 36.3%; Score 809.5; DB 2; Length 394;
Best Local Similarity 39.8%; Freq. No. 9.7e-39;
Matches 162; Conservative 66; Mismatches 134; Indels 45; Gaps 8;

Qy 1 MESDSTKISSISSSSSRPKYICTEGCDKANKRNSLLEQHRT-HSNDRPYKCTV 59
Db 13 LSRSSSVGSMSS-VAGGSVRSKTYLCEIDNCDKATRSLSLTHQNTVHLGKPKKC-- 69
Qy 60 DDCKAFPRKSHLETHIVSHSEKKPFCVCGKGVNSROHLKRIHTHTKSPKCTPENCO 119

Db 70 NCCESSFFKTKHLEHLYTHDERPFCYSCFCCKGLITRQOLRRHVTHTKSNCEYEGCN 129
Qy 120 EAFYGHOSLRHHILSVHEKTLTCKQCNKVFPPSKLAQHKLGHHGS--PAYQCHPGCF 177
Db 130 EEPYHQPOLRAHILVHLSQSLKCHECNKSFGQPYRLKRIAHGHPDVVNAIQCTFVSVC 189
Qy 178 KNPQTSVLOPHIKOSHPLKCPKCGCVGKGLSHMLSHDSTMIKIWTCDYCVG 237
Db 190 KSPFTWSALRLHVKNDDHRLKPCICSPCVGEGDGLNMHMKIDENILVSRNWKHCINDDQ 249
Qy 238 FAKKNEIYVHNIPHDGNIIPDLLKETEYKLENLDDQSKLNNLHELETEKLYEBDE 297
Db 250 FAKKNEIYVHNIPHDGNIIPDLLKETEYKLENLDDQSKLNNLHELETEKLYEBDE 293
Qy 298 DEEDSLDEKRSVDRSDMSAQRISKFTASLEGSKSVKLSNS--GKIKNCPKNCDRMF 356
Db 294 -----SPKKYKVAKTDMLAITETNNLRFPFGGDANTLLNTYGRKFRCCYSKCYRSF 347
Qy 357 SREYDLRRLKWHDDNLQRIESFLNS-----IEKEE 387
Db 348 KTEEKYNHIEKGRHIEQKLEKLETSQOLQVADIEPRQSTALEKEE 394

RESULT 6

TF3A_SCHPO STANDARD; PRT; 374 AA.
AC 09UTL5;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 05-JUL-2004 (Rel. 44, Last annotation update)
DE Transcription factor IIRA (Factor A) (TFIIRA).
GN Name=sfc2; ORFNames=SPAC144.09c;
OS Schizosaccharomyces pombe (Fission Yeast).
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC Schizosaccharomycetales; Schizosaccharomycetaceae;
OC Schizosaccharomycetes.
OX NCBI_TaxID=4896;
RN [1]
RP SEQUENCE FROM N.A. AND FUNCTION.
RC MEDLINE=22082319; PubMed=12087160; DOI=10.1093/nar/gk385;
RX Schuman D.B., Seizer D.R.;
RT "Identification and characterization of transcription factor IIRA from
Schizosaccharomyces pombe."
RL Nucleic Acids Res. 30:2772-2781(2002).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=972;
RX MEDLINE=21848401; PubMed=11859360; DOI=10.1038/nature724;
RX Wood V., Gwilliam R.G., Rajandream M.A., Lyne M., Lyne R., Stewart A.,
RA Sgouras J., Peat N., Hayles J., Baker S., Basham D., Bowman S.,
RA Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,
RA Collins M., Connor R., Cronin A., Davis P., Fellwell T., Fraser A.,
RA Gentles S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodgson G.,
RA Holroyd S., Hornaby T., Howarth S., Huckle E.J., Hunt S., Jagsels K.,
RA James K., Jones L., Jones M., Leacher S., McDonald S., McLean J.,
RA Mooney P., Moule S., Mungall K., Murphy L., Niblett D., Odell C.,
RA Oliver K., O'Neill S., Pearson D., Quail M.A., Rabinovitch B.,
RA Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S.,
RA Skelton J., Simmonds W., Squares R., Squares S., Stevens K.,
RA Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,
RA Woodward J., Volckaert G., Aert R., Robben J., Grymmondez B.,
RA Wellfens I., Vantermeir E., Rieger M., Schaefer M., Wellner-Auer S.,
RA Gabel C., Fuchs M., Fritz C., Holzer B., Moestl D., Hilbert H.,
RA Borzym K., Langer I., Beck A., Lehrach H., Reinhardt R., Pohl T.M.,
RA Eger P., Zimmermann W., Wedler H., Wambut L., Purnelle B.,
RA Goffeau A., Cadieu E., Dreano S., Gloux S., Lelaude V., Mottier S.,
RA Galibert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,
RA Lucas M., Rochet M., Galliardin C., Tallada V.A., Garçon A., Thode G.,
RA Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Bento J.,
RA Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Forsburg S.L.,
RA Cerutti L., Lowe T., McCombie W.R., Paulsen I., Potashkin J.,
RA Shpakovski G.V., Ussery D., Batteil B.G., Nurse P.;

RT "The genome sequence of Schizosaccharomyces pombe."
RL Nature 415:871-880(2002).
CC -!- FUNCTION: Is required for correct transcription of 5S RNA genes by
CC RNA polymerase III. Also binds the transcribed 5S RNA's. Initiates
CC transcription of the 5S ribosomal RNA gene.
CC -!- SUBCELLULAR LOCATION: Nuclear (By similarity).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AY091590; AM00046.1; -;
DR EMBL: AL132675; CAB59689.1; -;
DR PIR: T37676; T37676.
DR HSP: P07248; 1AR0.
DR GeneDB Sprobe: SPAC144.09c; -;
DR InterPro: IPR007087; Znf_C2H2.
DR Pfam: PF00096; Zf-C2H2; 10.
DR ProDom: PD000003; Znf_C2H2; 1.
DR SMART: SM00355; Znf_C2H2; 10.
DR PROSITE: PS00028; ZINC_FINGER_C2H2_1; 8.
DR PROSITE: PS50157; ZINC_FINGER_C2H2_2; 9.
DR DNA-binding; Metal-binding; Nuclear protein; Repeat; RNA-binding;
KM Transcription regulation; Zinc-finger.
FT ZN_FING 23 47 C2H2-type.
FT ZN_FING 53 77 C2H2-type.
FT ZN_FING 83 107 C2H2-type.
FT ZN_FING 113 138 C2H2-type.
FT ZN_FING 144 169 C2H2-type.
FT ZN_FING 204 226 C2H2-type.
FT ZN_FING 236 261 C2H2-type.
FT ZN_FING 267 291 C2H2-type (acyclic).
FT ZN_FING 349 374 C2H2-type.
SQ SEQUENCE 374 AA; 43851 MW; 7469C701FF08F6 CRC64;

Query Match 22.2%; Score 495; DB 1; Length 374;
Best local Similarity 32.8%; Pred. No. 9,7e-21;
Matches 123; Conservative 58; Mismatches 160; Indels 34; Gaps 13;

QY 15 SSSSSSPKRYICTYEGCDAYNRPRLLEQLRTHSDRYKCTVDDCDKAFPRKSLTET 74
DB 13 SKULRSKKLFHCPEYECGKYSRPSLLEQLRTHSDRYKCTVDDCDKAFPRKSLTET 72
QY 75 HIYSHSEKDPFHCVS--CGKGVNSROHLKXHEITH--TKSFKCTFENCQBAFYKQSLRH 130
DB 73 HKGCHTNVSPFSCHYDGDCAQFTYQDLERHIEVHRKPKYACTWEGDCDFSKHQQLRS 132
QY 131 HILSVHEKTL----TCQKNCNVFTRPSKLAQHLKHHGSPAYQCDHPC--FKNFQWS 184
DB 133 HISACHTHLPYPCTYDDCELRATKQKQLQNHVNBAREKILISVCPHESVGHGEFKWS 192
QY 185 VLOPHIKQSPKLCPCPCGKGVCKKLSHMLSHDSTMIKTYTCDYCV---GKFAK 240
DB 193 QLNHLHREAHVP--SCSICGRQFKTAHLRHHVVLH--QTLLEKTYHCHMEGCKSPFR 249
QY 241 KNEIIVENYLFHQGNIP---DDLKETEYK--KLENLLDQSKLNNLHELETEKLYAYEDE 296
DB 250 SSALKKHSIYIHGNNMFPHDCSGCTKFGYKMLQRLHERGT--CKNAHKPYINEGCIHGD 308
QY 297 EDEBDSIDEKRSVDVSDMS--AQRISKSFASLEGSKVSKLISNGKKINCPKNNCDR 354
DB 309 IEGVALIDQKEKELSSMLVSDVAKKILNEVTG--HGK-----EAREVSGSPFECNY 358
QY 355 MFSREYLLRRHLKMH 369
DB 359 RFRKLYDMHRLHNSH 373

Q6CZU4
ID Q6CZU4 PRELIMINARY; PRT; 655 AA.
AC Q6CZU4;
DT 25-OCT-2004 (TrEMBLrel. 28, Created)
DT 25-OCT-2004 (TrEMBLrel. 28, Last sequence update)
DT 25-OCT-2004 (TrEMBLrel. 28, Last annotation update)
DE Yarrowia lipolytica chromosome F of strain CUB99 of Yarrowia
DE lipolytica.
GN ORFNames=YAL10F05104g;
OS Yarrowia lipolytica CUB99.
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Dipodascaceae; Yarrowia.
OX NCBI_TaxId=284591;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CUB99;
RG Genolevures;
RA Dujon B., Sherman D., Fischer G., Durrens P., Casaregola S.,
RA Lafontaine I., de Montigny J., Marck C., Neugeglisse C., Talla E.,
RA Goffard N., Frangeul L., Aigle M., Anthonard V., Babour A., Barbe V.,
RA Barnay S., Blanchin S., Beckerich J.M., Beyne E., Bleykasten C.,
RA Boissaye A., Boyer J., Catolico L., Contandriou F., de Daruvar A.,
RA Despons L., Fabre B., Fairhead C., Ferry-Dumazet H., Gropi A.,
RA Hantreay F., Hennequin C., Jaumaux N., Joyet P., Kachouri R.,
RA Kerrest A., Kozul R., Lemaire M., Leguier I., Ma L., Muller H.,
RA Niclaud J.M., Nikolek M., Oztas S., Ozier-Kalogeropoulos O.,
RA Pellenz S., Potier S., Richard G.F., Straud M.L., Suleau A.,
RA Swenne D., Tekala F., Wesolowski-Louvel M., Westhof E., Wirth B.,
RA Zenlou-Meyer M., Zivanovic I., Bolotin-Fukuhara M., Thierry A.,
RA Bouchier C., Caudron B., Scarpelli C., Galliard C., Weisenbach J.,
RA Wincker P., Souciet J.L.;
RT "genome evolution in yeasts."
RL Nature 430:35-44(2004).
[2]
RN SEQUENCE FROM N.A.
RP STRAIN=CUB99;
RC Genoscope;
RA Submitted (JUL-2004) to the EMBL/GenBank/DBJ databases.
RL EMBL: CR382132; CAG77825.1; -;
DR GO: GO:0005634; C:nucleus; IEA.
DR GO: GO:0003676; P:nucleic acid binding; IEA.
DR GO: GO:0008270; F:zinc ion binding; IEA.
DR InterPro: IPR006025; Pept_M_Zn_BS.
DR InterPro: IPR007087; Znf_C2H2.
DR Pfam: PF00096; Zf-C2H2; 5.
DR ProDom: PD000003; Znf_C2H2; 1.
DR SMART: SM00355; Znf_C2H2; 10.
DR PROSITE: PS00028; ZINC_FINGER_C2H2_1; 7.
DR PROSITE: PS50157; ZINC_FINGER_C2H2_2; 3.
DR PROSITE: PS00142; ZINC_PROTEIN; UNKNOWN 1.
SQ SEQUENCE 655 AA; 74609 MW; 5FF764F312CF4882 CRC64;

Query Match 21.1%; Score 470.5; DB 2; Length 655;
Best local Similarity 28.9%; Pred. No. 4,6e-19;
Matches 111; Conservative 65; Mismatches 145; Indels 63; Gaps 11;

QY 21 RPKYICTYEGCDAYNRPRLLEQLRTHSDRYKCTVDDCDKAFPRKSLTETVSHS 80
DB 83 RFLAHVCPREGCDKAYSRPSLLEQLRTHSDRYKCTVDDCDKAFPRKSLTETVSHS 142
QY 81 EKKPFHCVSQGVNSROHLKXHEITHTSFKCTFENCQBAFYKQSLRHILSVH---E 137
DB 143 VAKRYHCSFCAKFNRQHLRHEVTHKSIKCPHDCSQCTEVQLEHCVADHATWK 202
QY 138 KTLTCQKNCNVFTRPSKLAQHLKHHGSPAYQCDHPCFKNFQWSVLQFHLKQSPK 197
DB 203 HQCTYEGCGRAIGSERQWQYHMERDHNQTPQYVCPHDECEVASQWIKQKHVSXAHNR 262
QY 198 KCPKCGKGVCKKGL--SSHMLSHDST-----MIKTYC--DYCDVG 236
DB 263 LCHHC-----EGIDPDVQSWYEHKSVGLSVLYRGGRTPGYROLMTCPESCGHQ 315
QY 237 KFAKNEIIVENYLFHQGNIPD--DLKETEY-----KLENLLDQSKLNNLHELE 286

```

Db      316 SF-DKOLYAKHYVSHNRKAPRDMFESVAVYDITPELIPFEKPEKLTIRGK-----R 365
Oy      287 TEKLKVEEDBEDSDLDKERSDVSFSDMSAGRSIKSFASLEGSVKLSNSG----- 342
Db      366 SERTSLSEGSGBEEYDLPSER-----PAKQGRKREKSPKPTAASIVDKITGAGYSOS 417

Oy      343 KXINCPKNNCDRMFSREYDLRRHL 366
Db      418 RTIPCVVETCMHFRNRKYLDRHV 441

RESULT 8
ID      06RSP3      PRELIMINARY;      PRT;      634 AA.
AC      06RSP3;
DT      05-JUL-2004 (TREMBLrel. 27, Created)
DT      05-JUL-2004 (TREMBLrel. 27, Last sequence update)
DT      05-JUL-2004 (TREMBLrel. 27, Last annotation update)
DE      KRAB-zinc finger protein.
GN      Name=Ic68;
OS      Mus musculus (Mouse).
OC      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC      Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX      NCBI_TaxID=10090;
RN      (1)
RP      SEQUENCE FROM N.A.
RC      STRAIN=12956/SVETac;
RA      Brattwalte M., Maeltz P., Dudekula D., Qian Y., Nagaraja R.;
RA      Submitted (DEC-2003) to the EMBL/GenBank/DBJ databases.
DR      EMBL: AY510701; AAR91692.1; -.
DR      HSSP: P08046; IAIG.
DR      GO: GO:0005634; C:nucleus; IEA.
DR      GO: GO:0003676; P:nucleic acid binding; IEA.
DR      GO: GO:0008270; F:zinc ion binding; IEA.
DR      GO: GO:0003355; P:regulation of transcription, DNA-dependent; IEA.
DR      InterPro: IPR001909; KRAB.
DR      InterPro: IPR007087; Znf_C2H2.
DR      InterPro: IPR007086; Znf_C2H2_sub.
DR      Pfam: PF01352; KRAB; 1.
DR      Pfam: PF00096; zf_C2H2; 16.
DR      PRINTS: PR00048; ZINC_FINGER.
DR      ProDom: PD000003; Znf_C2H2; 15.
DR      SMART: SM00349; KRAB; 1.
DR      SMART: SM00355; Znf_C2H2; 16.
DR      PROSITE: PS00805; KRAB; 1.
DR      PROSITE: PS00028; F:zinc ion binding; IEA.
DR      PROSITE: PS00157; ZINC_FINGER_C2H2_2; 16.
DR      PROSITE: PS00157; ZINC_FINGER_C2H2_2; 16.
SQ      SEQUENCE 634 AA; 73724 MW; CBE21DDE4520SD3D CRC64;

Query Match      18.5%; Score 411.5; DB 2; Length 634;
Best Local Similarity 32.8%; Pred. No. 1.1e-15;
Matches 117; Conservative 54; Mismatches 129; Indels 57; Gaps 19;

Oy      23 KYITCTYEGCDYAVNRPSSLEQHLKTHSNDPRYKCTVDDCYAFKPKSHLETHIVSHSK 82
Db      323 KPYKC--KECGAFYTGSSSLKGLNLTGKPKYC--EECDKSFTEKSTLTTHKRIHTGK 378

Oy      83 KPFHSGVCGKGVNRRHLKRHEITHT--KSPKCTFENCOEARYKIOSLR--HILSNHEKT 139
Db      379 KPYKCNICGKSFSCFNLTHTDILTGEPKPYC--KECGKAPRYMSLLKIHONLHPGEKP 436

Oy      140 LTCKOCNKVFTRPSPKLAQHLKHHGSSPAYQCDHPGCFNPOFWSVLOFH--IKQSHPKL 197
Db      437 YKCKEDDKFTLKSLRTHQIITGSKP--YKCNV--CGSFPOCTNLKTHQHLTGKERY 493

Oy      198 KCPKCGKGVGKKGKGLSSHMLSHDSTMIKWTCDYCVKFAKKNELVEHYNIIFDGNIP 257
Db      494 KCKEKGKSFHYMSLTKSHQNLHSGE--KPYCKEED--KSFTEKSTLTGQRI--HTG--- 545

Oy      258 DLLKTEVTKLE--NLLDQGSGLNLTGHLTEKLVKEDEDEDESDLDKESDVRSDSM 315
Db      546 -----KKLYKCNICDK--SFTWCASLKTTH--KFNHGER-----PY 577

```

```

Oy      316 SAORSIKSF--TASLEGSVKLSNSGSKI--NCPKNNCDRMFSREYDLRRHLKWH 369
Db      578 KKKCKGKSPQSLTKRHOETHEDKGTGKLYKC--NDCKRSYSYHNSFRHQKIH 632

RESULT 9
ID      06PEP4      PRELIMINARY;      PRT;      634 AA.
AC      06PEP4;
DT      05-JUL-2004 (TREMBLrel. 27, Created)
DT      05-JUL-2004 (TREMBLrel. 27, Last sequence update)
DT      05-JUL-2004 (TREMBLrel. 27, Last annotation update)
DE      Hypothetical protein A830058L05.
GN      Name=A830058L05;
OS      Mus musculus (Mouse).
OC      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC      Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX      NCBI_TaxID=10090;
RN      (1)
RP      SEQUENCE FROM N.A.
RC      STRAIN=FVB/N; TISSUE=Mammary tumor. C3;
RC      MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;
RA      Strusberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA      Klausner R.D., Collins F.S., Wagner L., Shennan C.M., Schuler G.D.,
RA      Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA      Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA      Datchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA      Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.B.,
RA      Brownstein M.J., Usdin T.B., Toshiyuki S., Carroll P., Prange C.,
RA      Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,
RA      Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA      Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA      Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA      Fahy J., Helton E., Kettelman M., Madan A., Rodriguez S., Sanchez A.,
RA      Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA      Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA      Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Buterfield Y.S.,
RA      Krzywinski M.I., Skalska U., Smalins D.E., Scherch A., Schein J.B.,
RA      Jones S.J., Maitra M.A.;
RT      "Generation and initial analysis of more than 15,000 full-length human
RT      and mouse cDNA sequences."
RT      Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).
RN      [2]
RP      SEQUENCE FROM N.A.
RC      STRAIN=FVB/N; TISSUE=Mammary tumor. C3;
RA      Strusberg R.;
RA      Submitted (SEP-2003) to the EMBL/GenBank/DBJ databases.
DR      EMBL: BC057947; AAH57947.1; -.
DR      HSSP: P08046; IAIG.
DR      GO: GO:0005634; C:nucleus; IEA.
DR      GO: GO:0003676; F:nucleic acid binding; IEA.
DR      GO: GO:0008270; F:zinc ion binding; IEA.
DR      GO: GO:0003355; P:regulation of transcription, DNA-dependent; IEA.
DR      InterPro: IPR001909; KRAB.
DR      InterPro: IPR007087; Znf_C2H2.
DR      InterPro: IPR007086; Znf_C2H2_sub.
DR      Pfam: PF01352; KRAB; 1.
DR      Pfam: PF00096; zf_C2H2; 16.
DR      PRINTS: PR00048; ZINC_FINGER.
DR      ProDom: PD000003; Znf_C2H2; 15.
DR      SMART: SM00349; KRAB; 1.
DR      SMART: SM00355; Znf_C2H2; 16.
DR      PROSITE: PS00805; KRAB; 1.
DR      PROSITE: PS00028; ZINC_FINGER_C2H2_1; 16.
DR      PROSITE: PS00157; ZINC_FINGER_C2H2_2; 16.
DR      PROSITE: PS00157; ZINC_FINGER_C2H2_2; 16.
SQ      SEQUENCE 634 AA; 73682 MW; E6F7E83E452DEF24 CRC64;

Query Match      18.4%; Score 410.5; DB 2; Length 634;
Best Local Similarity 32.8%; Pred. No. 1.2e-15;
Matches 117; Conservative 54; Mismatches 129; Indels 57; Gaps 19;

```

QY 23 KKYCTYEGCDKAYNRPSTLEOHRLRTHSNDRPYKCTVDDCKAFPRKSHLETHIVSHSEK 82
 DB 323 KPYKC--KECGKAFQTSSSLKXHLNLTGKKPYKC--EECDKSFTEKSTLTTHKRJHTGK 378
 QY 83 KPRHGVCGKGVNRSOHLKRNHETHT--KSPKCTFENCQEAQYKHOSLR-HHILSVHEKT 139
 DB 379 KPRKNCICGKSFNSCTNLKTHQTLHTGKEKPYKC--KECGKAFYMSLSLKHQNLHNGEKEK 436
 QY 140 LTCQCNKVPTRSKLAONHLKHHGSPAYOCDPHCFKQFWSVLYQFH--IKOSHPLK 197
 DB 437 YKCEKCDKFTLKSJLRTHQIHTGKEKPYKC--CGKSFNCTNLKTHQNLHTGKEKPY 493
 QY 198 KCPKCGKGVCGKGLSSHMLSHDSTWIKITWCDYGVGKFAKKNELVEHYNIFFHGNIP 257
 DB 494 KCEKCGKSFHYMSLSKHQNLHSGE---KPYKCECD-KSFTKSTLIKQRI-HTG--- 545
 QY 258 DDLKKTETVKKLE--NLDDGSKLNNLHLETEKLYVEDEDEDEBDSLEKSDVSDSM 315
 DB 546 -----KKLYKNCIDK--SFTWCASLKTHK-KFHTGKEK-----PY 577
 QY 316 SAQRSIKSF--TASLEGSKSVSKLISNGSKI--NCPKNNCDRMFSREYDLRRLKMH 369
 DB 578 KCRKCGKSPQSLTLKRKHQKHEDKHTGKEKLYKC--NDCDRSYSHSFRRRQKXH 632

RESULT 10
 DB 08B185 PRELIMINARY; PRT; 634 AA.
 AC 08B185;
 DT 01-MAR-2003 (TREMBLrel. 23, Created)
 DT 01-MAR-2003 (TREMBLrel. 23, Last sequence update)
 DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)
 DE Mus musculus 10 days neonate cortex cDNA, RIKEN full-length enriched library, clone:A830058L05 product:stimilar to DNA-BINDING PROTEIN.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathia; Muridae; Murinae; Mus.
 NC NCB1_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6J; TISSUE=Cortex;
 RX MEDLINE=99279253; PubMed=10349636; DOI=10.1016/S0076-6879(99)03004-9;
 RA Carninci P., Hayashizaki Y.;
 RT "High-efficiency full-length cDNA cloning.";
 RL Meth. Enzymol. 303:19-44(1999).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6J; TISSUE=Cortex;
 RX MEDLINE=21085660; PubMed=11217851; DOI=10.1038/35055500;
 RA RIKEN FANTOM Consortium;
 RT "Functional annotation of a full-length mouse cDNA collection.";
 RL Nature 409:685-690(2001).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6J; TISSUE=Cortex;
 RA The FANTOM Consortium,
 RA the RIKEN Genome Exploration Research Group Phase I & II Team;
 RT "Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs.";
 RL Nature 420:563-573(2002).
 RN [4]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6J; TISSUE=Cortex;
 RX MEDLINE=2049374; PubMed=11042159; DOI=10.1101/gr.145100;
 RA Carninci P., Shibata Y., Hayatsu N., Sugahara Y., Shihata K., Itoh M., Kono H., Okazaki Y., Muramatsu M., Hayashizaki Y.;
 RT "Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes.";
 RL Genome Res. 10:1617-1630(2000).
 RN [5]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6J; TISSUE=Cortex;
 RX MEDLINE=20530913; PubMed=11076861; DOI=10.1101/gr.152600;

RA Shibata K., Itoh M., Aizawa K., Nagaoka S., Sasaki N., Carninci P., Kono H., Akiyama Y., Nishi K., Katsunari T., Taahiro H., Itoh M., Sumi N., Ishii Y., Nakamura S., Hazama M., Nishino T., Harada A., Yamamoto R., Matsumoto H., Sakaguchi S., Ikegami T., Kashiwagi K., Fujisaki S., Inoue K., Togawa Y., Izawa M., Ohara E., Watanabe M., Yoneda Y., Ishikawa T., Ozawa K., Tanaka T., Matsuzaki S., Kawai J., Okazaki Y., Muramatsu M., Inoue Y., Kita A., Hayashizaki Y.;
 RT "RIKEN integrated sequence analysis (RISA) system-384-format sequencing pipeline with 384 multicapillary sequencer.";
 RL Genome Res. 10:1757-1771(2000).
 RN [6]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6J; TISSUE=Cortex;
 RA Adachi J., Aizawa K., Akimura T., Aikawa T., Bono H., Carninci P., Rukuda S., Furuta M., Hanagata T., Hara A., Hashizume W., Hayashida K., Hayatsu N., Hiramoto K., Hiraoka T., Hirozane T., Hori F., Imotani K., Ishii Y., Itoh M., Kagawa I., Kasukawa T., Katoh H., Kawai J., Kojima Y., Kondo S., Kono H., Kouda M., Koya S., Kurihara C., Matsuyama T., Miyazaki A., Murata M., Nakamura M., Nishi K., Nomura K., Numazaki R., Ohno M., Ohnato N., Okazaki Y., Saito R., Saitoh H., Sakai C., Sakai K., Sakazume N., Sano H., Sasaki D., Shibata K., Shingawa A., Shitaki T., Sogabe Y., Tagami M., Tagawa A., Takahashi F., Takaku-Akahira S., Takeda Y., Tanaka T., Tomaru A., Toya T., Yasunishi A., Muramatsu M., Hayashizaki Y.;
 RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AK043955; BACJ1714.1; -.
 DR HSSP; P08045; 1ZNF.
 DR GO; GO:0005634; C:nucleus; IEA.
 DR GO; GO:0003677; F:DNA binding; IEA.
 DR GO; GO:0008270; F:zinc ion binding; IEA.
 DR GO; GO:0006355; P:regulation of transcription, DNA-dependent; IEA.
 DR InterPro; IPR001909; KRAB.
 DR InterPro; IPR007087; ZnF_C2H2.
 DR InterPro; IPR007086; ZnF_C2H2_sub.
 DR Pfam; PF01352; KRAB; 1.
 DR PRINTS; PR00096; zf_C2H2; 16.
 DR PROSITE; PRO0048; ZINC_FINGER.
 DR SMART; SM00349; KRAB; 1.
 DR SMART; SM00355; ZnF_C2H2; 16.
 DR PROSITE; PS00805; KRAB; 1.
 DR PROSITE; PS00028; ZINC_FINGER_C2H2_1; 16.
 DR PROSITE; PS00157; ZINC_FINGER_C2H2_2; 16.
 KW DNA-binding.
 SQ SEQUENCE 634 AA; 73668 MW; AA7DEB33D87508BA CRC64;
 Query Match 18.4%; Score 410.5; DB 2; Length 634;
 Best Local Similarity 32.8%; Pred. No. 1,2e-15;
 Matches 117; Conservative 54; Mismatches 129; Indels 57; Gaps 19;

QY 23 KKYCTYEGCDKAYNRPSTLEOHRLRTHSNDRPYKCTVDDCKAFPRKSHLETHIVSHSEK 82
 DB 323 KPYKC--KECGKAFQTSSSLKXHLNLTGKKPYKC--EECDKSFTEKSTLTTHKRJHTGK 378
 QY 83 KPRHGVCGKGVNRSOHLKRNHETHT--KSPKCTFENCQEAQYKHOSLR-HHILSVHEKT 139
 DB 379 KPRKNCICGKSFNSCTNLKTHQTLHTGKEKPYKC--KECGKAFYMSLSLKHQNLHNGEKEK 436
 QY 140 LTCQCNKVPTRSKLAONHLKHHGSPAYOCDPHCFKQFWSVLYQFH--IKOSHPLK 197
 DB 437 YKCEKCDKFTLKSJLRTHQIHTGKEKPYKC--CGKSFNCTNLKTHQNLHTGKEKPY 493
 QY 198 KCPKCGKGVCGKGLSSHMLSHDSTWIKITWCDYGVGKFAKKNELVEHYNIFFHGNIP 257
 DB 494 KCEKCGKSFHYMSLSKHQNLHSGE---KPYKCECD-KSFTKSTLIKQRI-HTG--- 545
 QY 258 DDLKKTETVKKLE--NLDDGSKLNNLHLETEKLYVEDEDEDEBDSLEKSDVSDSM 315
 DB 546 -----KKLYKNCIDK--SFTWCASLKTHK-KFHTGKEK-----PY 577
 QY 316 SAQRSIKSF--TASLEGSKSVSKLISNGSKI--NCPKNNCDRMFSREYDLRRLKMH 369
 DB 578 KCRKCGKSPQSLTLKRKHQKHEDKHTGKEKLYKC--NDCDRSYSHSFRRRQKXH 632

```
RESULT 11
ID 07RY67 PRELIMINARY; PRT; 611 AA.
AC 07RY67;
DT 01-MAR-2004 (TrEMBLrel. 26, Created)
DT 01-MAR-2004 (TrEMBLrel. 26, Last sequence update)
DE 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE Hypothetical protein.
OS Name=NCU00038.1;
OS Neurospora crassa.
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
OC Sordariomycetidae; Sordariales; Sordariaceae; Neurospora.
OX NCBI_TaxId=5141;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=OR74A;
RA Galagan J.E., Calvo S.E., Borkovich K.A., Selker E.U., Read N.D.,
RA Jaffe D., Fitzhugh W., Ma L.-J., Smirnov S., Purcell S., Rehan B.,
RA Ekins T., Engels R., Wang S., Nielsen C.B., Butler J., Endrizzi M.,
RA Qui D., Ianakiev P., Pedersen D., Nelson M., Washburne M.,
RA Seltrenitoff C.P., Kinsey J.A., Braun E.L., Zelter A., Schlute U.,
RA Kolbe G.O., Jedd G., Mewes W., Staben C., Marcotte E., Greenberg D.,
RA Roy A., Foley K., Naylor J., Thomann N., Barrett R., Gierre S.,
RA Kamal M., Kamysaselis M., Mauceli E., Bielke C., Rudd S., Fishman D.,
RA Krystofova S., Rasmussen C., Metzberg R.L., Perkins D.D., Kroken S.,
RA Cogoni C., Macino G., Catchside D., Li W., Pratt R.J., Osmari S.A.,
RA Desouza C.C., Glass L., Orbach M.J., Berglund J., Voelker R.,
RA Yarden O., Plamann M., Seiler S., Dunlap J., Radford A., Aramayo R.,
RA Nativy D., Alex L.A., Mannhaupt G., Ebbole D.O., Freitag M.,
RA Paulsen I., Sachs M.S., Lander E.S., Nussbaum C., Birren B.,
RT "The Genome Sequence of the Filamentous Fungus Neurospora crassa."
RL Nature 0:0-0(2003).
-1- CAUTION: The sequence shown here is derived from an
EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is
preliminary data.
CC EMBL; AABX01000729; EAA27789.1; -.
DR GO; GO:0005634; C:nucleus; IEA.
DR GO; GO:0003676; F:nucleic acid binding; IEA.
DR GO; GO:0008270; F:zinc ion binding; IEA.
DR InterPro; IPR007087; Znf_C2H2.
DR Pfam; PF00096; Zf-C2H2; 7.
DR PROSITE; PS00028; ZINC_FINGER_C2H2_1; 7.
DR PROSITE; PS50157; ZINC_FINGER_C2H2_2; 7.
KW Hypothetical protein.
SQ
SEQUENCE 611 AA; 67682 MW; 99584EAS410D8932 CRC64;

Query Match 18.3%; Score 407; DB 2; Length 611;
Best Local Similarity 24.9%; Pred. No. 1.8e-15;
Matches 111; Conservative 80; Mismatches 146; Indels 108; Gaps 17;

QY 1 MSESDETKSISSISSSSSRP-----KKYICVEGCDKAYNRPSSLEQHLRPHNSD 52
DB 91 LTHADPTTVAATSIINTTTTRASRRPSPDLKTRCTFACCTKFNPAALAAHLKSHND 150
QY 53 RPYKTVDDCDKAFPRKSHLETHIV-SHSEKPFHCSV--CGKGVNSRQHLKHEHTT-- 107
DB 151 RPFKCYDDCDKDYLEEKLSQHIKSHNDRKYCPREPGCKGSFYTNRLRLHVLVHGS 210
QY 108 TYSFKCT-FENCQEAFFYKQSLRHILSVHEKTLTKCKCNKVFTPRSKLAQHLKHHGS 166
DB 211 ADHYRCRGYGDCCVQSFRRKQTLQRIHRTVHL-----GK 243
QY 167 PAYQCDHPCGCFKNFQVWSVLQFIHQSHRKLK--CPKC-----GKCGVCKGK-----L 212
DB 244 SAPIPCNDGCDAGFDASALRRRVEHEHDLKFWCECAEBGDDSGRRVGFPTMLL 303
QY 213 SSHMLSHDSTMIKIWTCDYCVGKPAKKNLEVEHYNIPDGLKVEVKLENTL 272
DB 304 QAHMKKEHN-----CAFCGV-RCGTQSNMRRHVELHSAKYVD--RKTIACWEGC 353
QY 273 LDGSKLNLHLELETEKLKVEEDER--DEDSLDKRSDDVNSDSMSAQRSTKSTASIE 329
```

```
DB 354 DKKFTRVSNLN---THIKSAHEGHRFCVCGQDTDTYDAKYSLEIDWMPAEBCGCGFTTRVK 410
QY 330 GSKSVSKLISNGKK-----INCPKNKNCBMPFSEBYDL 362
DB 411 LEEHY--LHVHGKGRPRKLYVPSMWAAQQAQQAALDTRDLACPACTARFIRHDL 468
QY 363 RRLHKWHDNIQRISFINSIEKEE 387
DB 469 DRHIONDHGN-----NGIKKEE 485

RESULT 12
ID 08N703 PRELIMINARY; PRT; 404 AA.
AC 08N703;
DT 01-OCT-2002 (TrEMBLrel. 22, Created)
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DE 01-OCT-2002 (TrEMBLrel. 24, Last annotation update)
DE Hypothetical protein FLJ40479.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
OX NCBI_TaxId=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Testis;
RX PubMed=14702039; DOI=10.1038/ng1285;
RA Ota T., Suzuki Y., Nishikawa T., Otsuki T., Sugiyama T., Irie R.,
RA Wakamatsu A., Hayashi K., Sato H., Nagai K., Kimura K., Makita H.,
RA Sekine M., Ohsaishi M., Nishi T., Shibahara T., Tanaka T., Ishii S.,
RA Yamamoto J., Saito K., Kawai Y., Isono Y., Nakamura Y., Nagahara K.,
RA Murakami K., Yasuda T., Iwayanagi T., Wagatsuma M., Shiratori A.,
RA Sudo H., Hosoiri T., Kaku Y., Kodaira H., Kondo H., Sugawara M.,
RA Takahashi M., Kanda K., Yokoi T., Furuya T., Kikkawa E., Omura Y.,
RA Abe K., Kamihara K., Katsuta N., Sato K., Tanikawa M., Yamazaki M.,
RA Ninomiya K., Ishibashi T., Yamashita H., Murakawa K., Fujimori K.,
RA Tanai H., Kimachi M., Watanabe M., Hirooka S., Chida Y., Ishida S.,
RA Ono Y., Takiguchi S., Watanabe S., Yosida M., Houchi T., Kusano J.,
RA Kanohori K., Takahashi-Fujii A., Hara H., Tanase T., Nomura Y.,
RA Toguya S., Konoji H., Hara R., Takeuchi K., Arita M., Imose N.,
RA Musashino K., Yuuki H., Oshima A., Sagaki N., Naitouka S.,
RA Yoshikawa Y., Matsunawa H., Ichihara T., Shiohata N., Sano S.,
RA Moriya S., Momiyama H., Satoh N., Takami S., Terashima Y., Suzuki O.,
RA Nakagawa S., Senoh A., Mizoguchi H., Goto Y., Shimizu F., Wakebe H.,
RA Hishigaki H., Watanabe T., Sugiyama A., Takemoto M., Kawakami B.,
RA Yamazaki M., Watanabe K., Kumagai A., Itakura S., Fukuzumi Y.,
RA Fujimori Y., Komiyama M., Tashiro H., Tanigami A., Fujitawa T.,
RA Ono T., Yamada K., Fujii Y., Ozaki K., Hiro M., Omori Y.,
RA Kawabata A., Hikiji T., Kobatake N., Inagaki H., Ikema Y., Okamoto S.,
RA Okitani R., Kawakami T., Noguchi S., Itoh T., Shigeta K., Senba T.,
RA Matsumura K., Nakajima Y., Mizuno T., Morinaga M., Sasaki M.,
RA Togeishi T., Oyama M., Hata H., Watanabe M., Komatsu T.,
RA Mizushima-Sugano J., Satoh T., Shirai Y., Takahashi Y., Nakagawa K.,
RA Okumura K., Nagase T., Nomura N., Kikuchi H., Masuno Y., Yamashita R.,
RA Nakai K., Yada T., Nakamura Y., Ohara O., Isogai T., Sugano S.;
RT "Complete sequencing and characterization of 21,243 full-length human
cDNAs."
RL Nat. Genet. 36:40-45(2004).
DR EMBL; AK097798; BAC05174.1; -.
DR HSSP; P08047; ISP2.
DR GO; GO:0005634; C:nucleus; IEA.
DR GO; GO:0003676; F:nucleic acid binding; IEA.
DR GO; GO:0008270; F:zinc ion binding; IEA.
DR InterPro; IPR007087; Znf_C2H2.
DR InterPro; IPR007086; Znf_C2H2_sub.
DR Pfam; PF00096; Zf-C2H2; 14.
DR PRINTS; PR00048; ZINC_FINGER.
DR PRODOM; PD000003; Znf_C2H2; 14.
DR SMART; SM00355; Znf_C2H2; 14.
DR PROSITE; PS00028; ZINC_FINGER_C2H2_1; 12.
DR PROSITE; PS50157; ZINC_FINGER_C2H2_2; 14.
SQ
SEQUENCE 404 AA; 45923 MW; B1B35F62P5DF0007 CRC64;
```


	Query Match	17.9%	Score 400;	DB 2;	Length 404;	
	Best Local Similarity	33.9%;	Pred. No. 2.9e-15;			
	Matches 121;	Conservative 39;	Mismatches 163;	Indels 34;	Gaps 19;	
Oy						
Db						
Oy						
Db						
Oy						
Db						
Oy						
Db						
Oy						
Db						
Oy						
Db						
Oy						
Db						
Oy						
Db						
Oy						
Db						
Oy						
Db						
Oy						
Db						
Oy						
Db						
Oy						
Db						
Oy						
Db						
Oy						
Db						
Oy						
Db						
Oy						
Db						
Oy						
Db						
Oy						
Db						
Oy						
Db						
Oy						
Db						
Oy						
Db						
Oy						
Db						
Oy						
Db						
Oy						
Db						
Oy						
Db						
Oy						
Db						
Oy						
Db						
Oy						
Db						
Oy						
Db						

RT	"Generation and initial analysis of more than 15,000 full-length human
RL	and mouse cDNA sequences."
RN	Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RP	[3]
RA	SEQUENCE OF 38-190 FROM N.A.
RX	MEDLINE=91219421; PubMed=2023909;
RY	Bellefroid E.J., Poncellet D.A., Lecocq P.J., Revelant O.,
SA	Mattiel J.U.;
RT	"The evolutionarily conserved Kruppel-associated box domain defines a
RT	subset of eukaryotic multifingered proteins."
RL	Proc. Natl. Acad. Sci. U.S.A. 88:3608-3612(1991).
RP	[4]
RN	SEQUENCE OF 476-531 FROM N.A.
RC	TISSUE=Lymphoid;
RX	MEDLINE=91145339; PubMed=2288909;
RA	Thiesen H.-J.;
RT	"Multiple genes encoding zinc finger domains are expressed in human T
RL	cells."
CC	New Biol. 2:363-374(1990).
CC	-I FUNCTION: May be involved in transcriptional regulation.
CC	-I SUBCELLULAR LOCATION: Nuclear (Probable).
CC	-I TISSUE SPECIFICITY: T and B cell lines.
CC	-I SIMILARITY: Belongs to the kruepel C2H2-type zinc-finger protein
CC	family.
CC	-I SIMILARITY: Contains 1 KRAB domain.
CC	-----
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC	the European Bioinformatics Institute. There are no restrictions on its
CC	use by non-profit institutions as long as its content is in no way
CC	modified and this statement is not removed. Usage by and for commercial
CC	entities requires a license agreement (See http://www.isb.ch/announce/
CC	or send an email to license@isb-sib.ch).
CC	-----
DR	EMBL; X59244; CA41932.1; -.
DR	EMBL; BC006528; AAH06528.2; ALT_INIT.
DR	EMBL; M61869; AAA58674.1; -.
DR	EMBL; X52358; CAA3584.1; -.
DR	PIR; S26823; S26823.
DR	HSSP; P08048; 7ZNF.
DR	TRANSFAC; T04986; -.
DR	Genew; HGNC:13109; ZNF43.
DR	MIM; 603972; -.
DR	GO; GO:0003677; F: DNA binding; TAS.
DR	InterPro; IPR001909; KRAB.
DR	InterPro; IPR007087; ZnF_C2H2.
DR	InterPro; IPR007086; ZnF_C2H2_sub.
DR	Pfam; PF01352; KRAB_1.
DR	Pfam; PF00096; zF-C2H2; 21.
DR	PRINTS; PR00048; ZNCFINGER.
DR	Prodom; PD000003; ZnF_C2H2; 22.
DR	SMART; SM00349; KRAB_1.
DR	SMART; SM00355; ZnF_C2H2; 22.
DR	PROSITE; PS50805; KRAB; 1.
DR	PROSITE; PS50028; ZINC_FINGER_C2H2_1; 19.
DR	PROSITE; PS50157; ZINC_FINGER_C2H2_2; 22.
KM	DNA-binding; Metal-binding; Nuclear protein; Repeat;
KW	Transcription regulation; Zinc-finger.
FT	DOMAIN
FT	1
FT	69
FT	ZN_FING
FT	168
FT	190
FT	218
FT	C2H2-type.
FT	(degenerate).
FT	ZN_FING
FT	224
FT	C2H2-type (degenerate).
FT	252
FT	274
FT	C2H2-type (degenerate).
FT	280
FT	302
FT	C2H2-type.
FT	ZN_FING
FT	308
FT	330
FT	C2H2-type.
FT	336
FT	C2H2-type.
FT	364
FT	C2H2-type.
FT	392
FT	C2H2-type.
FT	414
FT	C2H2-type.
FT	420
FT	C2H2-type.
FT	448
FT	C2H2-type.
FT	470
FT	C2H2-type.
FT	498
FT	C2H2-type.
FT	504
FT	C2H2-type.
FT	532
FT	C2H2-type.
FT	554
FT	C2H2-type.

RESULT 15

ID	O43693	PRELIMINARY;	PRT;	395 AA.
AC	O43693;			
DT	01-JUN-1998 (TrEMBLrel. 06, Created)			
DT	01-JUN-1998 (TrEMBLrel. 06, Last sequence update)			
DT	01-MAR-2004 (TrEMBLrel. 26, Last annotation update)			
DE	Zinc-finger protein (Fragment).			
GN	Name=ZFS-1;			
OS	Homo sapiens (Human).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.			
OX	NCBI_TaxID=9606;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	TISSUE=Seminoma.			
RX	MEDLINE=98069676; PubMed=9406578; DOI=10.1016/S0165-4608(97)00004-6;			
RA	Ogawa T., Poncellet D., Kinoshita Y., Noce T., Takeda M., Kawamoto K.,			
RA	Udagawa K., Iecocq P., Martine J., Martial J., Hosaka M.;			
RT	"Enhanced expression in seminoma of human zinc finger genes located on			
RT	chromosome 19."			
RL	Cancer Genet. Cytogenet. 100:36-42(1998).			
DR	EMBL; D70831; BAA24050.1; -.			
DR	HSSP; P08153; 1ZFD.			
DR	GO; GO:0005634; C:nucleus; IEA.			
DR	GO; GO:0003676; F:nucleic acid binding; IEA.			
DR	GO; GO:0008270; F:zinc ion binding; IEA.			
DR	InterPro; IPR006025; Pept_M_Zn_BS.			
DR	InterPro; IPR007087; ZnF_C2H2.			
DR	InterPro; IPR007086; ZnF_C2H2_sub.			
DR	Pfam; PF00096; zf-C2H2; 13.			
DR	PRINTS; PR00048; ZINC_FINGER.			
DR	ProDom; PD000003; ZnF_C2H2; 12.			
DR	SMART; SM00355; ZnF_C2H2; 13.			
DR	PROSITE; PS00028; ZINC_FINGER_C2H2_1; 13.			
DR	PROSITE; PS00157; ZINC_FINGER_C2H2_2; 13.			
DR	PROSITE; PS00142; ZINC_PROTEASE; UNKNOWN_1.			
KW	Zinc-finger.			
KW	NON_TER			
FT	NON_TER 1 1			
SQ	SEQUENCE 395 AA; 45291 MW; D4E06B54DCE0BF9B CRC64;			

Query Match 17.8%; Score 397.5; DB 2; Length 395;
 Best Local Similarity 35.9%; Pred. No. 4e-15;
 Matches 97; Conservative 35; Mismatches 87; Indels 51; Gaps 14;

QY	23	KXYICTYEGCDKAYNRPSSLLBOHLRTHSNDPRPYKCTVDDCKAFPRKSHLETHIVSHSEK 82
DB	133	KPYKC--KECGKAFNRSSLTTRKIHTEGKPYKC--EECGKALQSSHLTHKIHTEG 188
QY	83	KPRHGSVCCGVNVRSHQIKRHEITHT--KSFKCTFENCOEAF----- 122
DB	189	KPYKCKKCGAFQSAHLTTHVEVHTEGKPYKC--EKCGKAFNHFSLTHTKIHTEGKP 246
QY	123	-----YKHOS--LRHHILSVHEKTLCTKCOCNKVFTRPSKLAOHKLKHGSPAYOC 171
DB	247	YKCKEGKAPKHSSTLTAKKIHTEGKPYKCKEGKAFNQSSTLTENK--KIHTGKPYEC 305
QY	172	DHPCGFKNFQTVSWLQFHIKQSHPK--LKCPRCKGKGVGKGLSSHMLSHDSTWIKIW 228
DB	306	EE--CGKAFNQSNTLRH--KKSHTGKPYKCEGKALNRPSTLTIHKIH---TGEKPY 359
QY	229	TCDYCVGK--FAKKNELVEHYNI FHDGNIP 257
DB	360	KCEEC--GKAFNQSSTLTAKKIHTEGKPY 386

Search completed: February 8, 2005, 15:22:03
 Job time : 183 secs

This Page Blank (uspto)